University of Glasgow

Structural and biophysical characterisation of the translocation and assembly module (TAM) complex from *Escherichia coli*

A thesis submitted to the University of Glasgow for the degree of Doctor of Philosophy

Inokentijs Josts BSc (Hons), MRes

Submitted October 2014

Institute of Infection, Immunity and Inflammation

College of Medical, Veterinary and Life Sciences

University of Glasgow

Acknowledgements

First and foremost, I would like to thank my supervisors Dan and Olwyn who have advised and supported me throughout my whole PhD, in and out of the lab. Additionally, I would like to extend my biggest thanks to the present and past members of the Walker lab. I am grateful to have spent the past four years of my life working alongside you all. You guys were like family to me. I will miss you all. You rock! Additionally, I would like to thank everyone on Level 2 for helping me during my PhD.
A special thanks goes to Brian for helping me with all the NMR work in and out of this thesis, Sharon for all the help with CD and for letting me use her ITC machine practically day and night, Aleks for all the help with X-rays

Additionally I would like to thank Olwyn, Darren and the University of Glasgow for giving me this opportunity and the Wellcome Trust for a very generous financial support.

Finally a big thank you goes to my girlfriend Kate and my family who have supported me throughout my PhD and beyond, I am very grateful for everything.

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or at any other institution.

Inokentijs Josts

October 2014

Abstract

Bacterial proteins destined to function extracellularly must efficiently cross two layers of bacterial cell membrane. Numerous transport systems are employed to ensure the delivery of these proteins to their final destination in a fully functional state. Autotransporters are outer membrane proteins that act as virulence factors in the extracellular milieu by mediating contact with the host and through manipulation of host defences. The translocation assembly module (TAM) complex is a cell envelope-spanning complex produced by Gram-negative bacteria that facilitates efficient secretion of autotransporters across the outer membrane. This complex consists of a member of the Omp85 superfamily of proteins, the outer membrane protein TamA, which in turn comprises an integral membrane β-barrel and three soluble periplasmic polypeptide transport (POTRA) domains, and a large inner membrane anchored protein TamB.

In this thesis, the binding determinants of TAM complex formation are delineated using biophysical and structural techniques. Association between TamA and TamB is mediated mostly through the interaction of the central polypeptide transport domain of TamA and the C-terminal region of the DUF490 domain of TamB. It is also demonstrated that TamB, in addition to interacting with TamA, forms a complex with the N-terminal region of the passenger domain of an autotransporter substrate, Ag43, that exists in an unfolded, extended conformation *in vitro*. Additionally, the TamA and Ag43 binding epitopes of TamB are overlapping but not identical. These data suggest that TamB acts as a chaperone that can exist in equilibrium between a TamA-bound form and an Ag43-bound form, and the formation of these complexes acts to deliver the autotransporter passenger domain to TamA.

Weak, transient interactions between the unstructured Ag43 passenger domain polypeptide and one of the periplasmic POTRA domains of TamA are also demonstrated. These transient interactions presumably act to guide the unfolded autotransporter polypeptide into the TamA barrel for efficient secretion across the outer membrane.

Table of Contents

List of Tables

List of Figures

Chapter 1

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Abbreviations

ΔH	Change in enthalpy
ΔS	Change in entropy
ABC	ATP-binding cassette
Ag43	Antigen 43
Aro	Aromatic
Amp	Ampicillin
ATOM	Archaic translocase of outer mitochondrial membrane
AT	Autotransporter
AUC	Analytical ultracentrifugation
ATP	Adenosine triphosphate
BAM	β-barrel assembly machine
CD	Circular dichroism
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CSP	Chemical shift perturbation
Da	Dalton
DMSO	Dimethyl sulfoxide
dH_20	Distilled water
DNA	Deoxyribonucleic acid
DMD	Discrete molecular dynamics
DUF	Domain of unknown function
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EOM	Ensemble optimisation method
ESI MS	Electrospray ionisation mass spectrometry
FHA	Filamentous haemagglutinin
FkpA	FKBP-type peptidyl-prolyl *cis-trans* isomerase
HRP	Horseradish peroxidase
HSQC	Heteronuclear single quantum coherence
IB	Inclusion body
IDP	Intrinsically disordered protein
IM	Inner membrane
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
Kan	Kanamycin
kb	Kilobase pair

K_d	Equilibrium dissociation constant
k_{on}	Association rate constant
k_{off}	Dissociation rate constant
kDa	Kilodalton
LB	Lysogenic Broth
LDAO	Lauryldimethylamine-oxide
LPS	Lipopolysaccharide
MTSSL	(1-Oxyl-2,2,5,5-tetramethylpyrroline-3-methyl) methanethiosulfonate
M_W	Molecular mass
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser effect spectroscopy
OD_{600}	Optical density at 600 nm
OMP	Outer membrane protein
OM	Outer membrane
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
pI	Isoelectric point
pmf	Proton motive force
POTRA	Polypeptide-translocation associated
PPIase	Peptidyl-prolyl isomerase
p.p.m.	Parts per million
PRE	Paramagnetic relaxation enhancement
r.m.s.d.	Root-mean-squared deviation
r.p.m.	Revolutions per minute
R_g	Radius of gyration
SAM	Sorting and assembly machinery
SAXS	Small angle X-ray scattering
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SE	Sedimentation equilibrium
SEC	Size exclusion chromatography
Skp	Seventeen kilodalton protein
SOMO	Solution modeller
SP	Signal peptide
SRP	Signal recognition particle
SurA	Survival protein A
SV	Sedimentation velocity
TAM	Translocation and assembly module
TBE	Tris Borate EDTA
TCEP	tris(2-carboxyethyl)phosphine

TOCSY	Total correlation spectroscopy
TOM	Translocase of the outer mitochondrial membrane
TPSS	Two-partner secretion system
UV	Ultraviolet
v/v	Volume per volume
WT	Wild-type
w/v	Weight per volume

1 Introduction

1.1 Bacterial cell envelope

The cell envelope of Gram-negative bacteria, such as *E. coli*, consists of two layers of membrane separated by the periplasmic space. The inner membrane is made up of a bilayer of symmetrically distributed phospholipids along with integral membrane proteins and peripheral lipoproteins anchored to it. These proteins generally possess housekeeping functions and are involved in energy production, cell division, signal transduction, lipid synthesis, as well as protein transport and secretion (1). The main structural anchors of integral inner membrane proteins consist of hydrophobic α-helices which are embedded into the lipid bilayer during secretion via the Sec system (2). Peripheral lipoproteins are embedded into the membrane via acyl chains covalently linked to the N-terminal cysteine amino acid (3). The periplasm harbours the peptidoglycan layer, which is essential for cell integrity and maintenance of turgor pressure. The peptidoglycan consists of polymers of N-acetylglucosamine and N-acetylmuramic acid cross-linked by short peptides forming a continuous mesh around the inner membrane (4). Numerous periplasmic proteins found in the periplasmic space play a role in peptidoglycan recycling and maintenance, for example during cell division or as an adaptive response to osmotic changes in the environment (5). The outermost membrane is a unique structure that is crucial to bacterial survival. It forms the primary source of solute and nutrient transport into and out of the cell, at the same time acting as a protective layer since it is impermeable to large hydrophobic molecules and toxins that would otherwise be harmful to the cell (6). In contrast to the inner membrane, the outer membrane is asymmetric with the inner leaflet made up of phospholipids and the outer leaflet consisting of lipopolysaccharide. Moreover, proteins found embedded into the outer membrane differ structurally from inner membrane proteins inasmuch as they possess β-barrel domains, usually consisting of 12, 16 or 22 anti-parallel β-strands joined by loops, with additional soluble domains facing the periplasm or the cell exterior. Integral outer membrane proteins (OMPs) serve a variety of functions from solute and nutrient transport, e.g. porin proteins PhoE and OmpF, to aggregation, adhesion and attachment, Ag43, and secretion, e.g. autotransporter proteins (7,8). A significant

number of these OMPs, in particular autotransporters (ATs), act as virulence factors during bacterial infection.

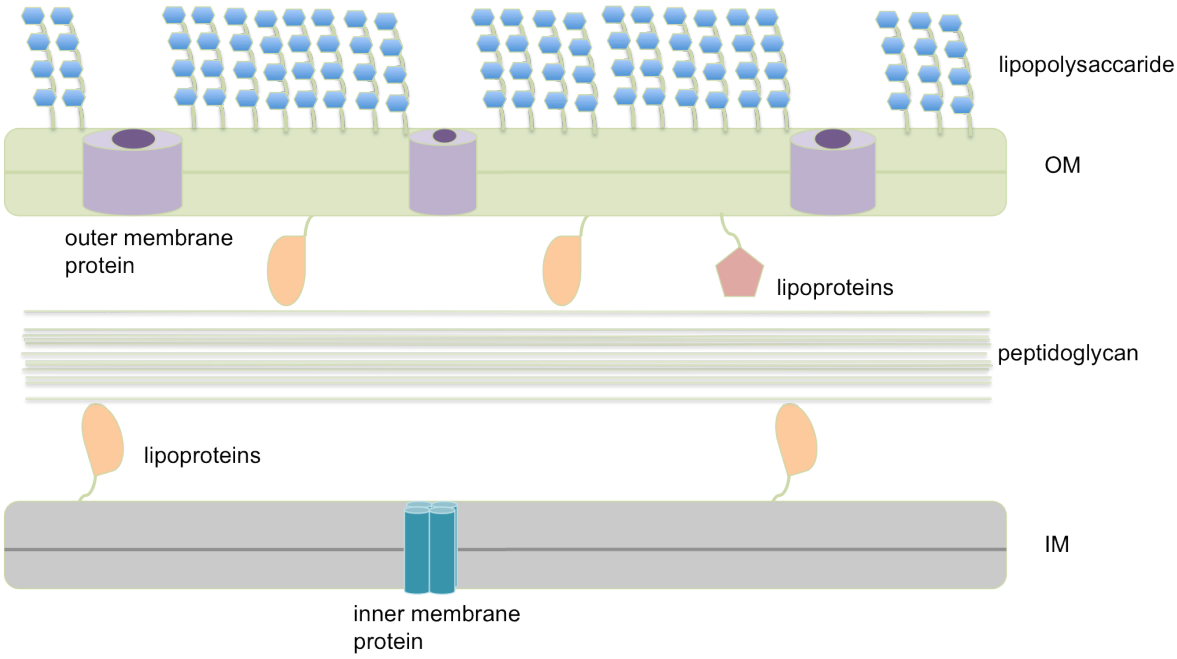


Figure 1-1 Overview of the cell envelope of Gram-negative bacteria. Bacterial cells are surrounded by two membranous layers separated by the periplasmic space harbouring the cell wall (sacculus). Proteins found in the inner membrane possess one or more α-helical transmembrane-spanning motifs (integral membrane proteins) or covalently attached acyl chains (lipoproteins). The IM bilayer is symmetric with regards to the phospholipid content. The OM bilayer is asymmetric, with the inner leaflet possessing phospholipids and the outer leaflet mostly made of phospholipids and lipopolysaccharide. Integral OMPs span the lipid layer by forming β-barrel structures.

1.1.1 *Bacterial outer membrane*

The OM bilayer forms a physical barrier between the bacterial cell and the environment, obstructing chemical exchange between the periplasm and the extracellular side. However, due to the presence of integral OM porins embedded into the lipid bilayer, simple solutes of molecular mass up to 600 Da can be exchanged by passive diffusion. OM integrity is additionally supported by OMPs and lipoproteins anchored to the periplasmic peptidoglycan layer. The predominant OM phospholipids are phosphatidylethanolamine (PE), phostatidylglycerol and cardiolipin (9). The outer leaflet is rich in lipopolysaccharide

(LPS), a tripartite assembly of lipid A, coupled to an oligosaccharide core region and extended branches of oligosaccharide repeats (O-antigen). LPS, incorporated in the OM bilayer, provides protection to the bacterial cell against harmful hydrophobic molecules as well as assistance in the attachment of bacterial cells to surfaces (10,11). LPS serves as one of the primary activators of the innate immune system in animals and plants infected with virulent bacteria and is therefore malleable in terms of diversity and modifiability in order to escape the immune response within the host (12). LPS molecules are synthesised within the inner leaflet of the inner membrane (IM) and translocated across the IM by an ATP-binding cassette (ABC) transporter MsbA to the periplasmic side (13). LPS molecules then cross the periplasmic space with the aid of LptA and are delivered to the OM multi-protein complex consisting of LptD, an integral OMP and LptE, a lipoprotein tightly associated with LptD (14,15). Recently, the structure of the LptDE complex was solved by X-ray crystallography to reveal a 26-stranded β-barrel that suggests incorporation of translocating LPS molecules into the lipid bilayer through a lateral opening mechanism of the LptD barrel (16). LPS molecules have additionally been implicated in assisting the folding and assembly of integral OMPs; however, the precise function of LPS in the assembly of OMPs is currently unknown (17).

Permeability of the OM mainly arises due to the presence of numerous porin proteins embedded into the bilayer. Porins form permeable (sometimes multimeric) channels due to their β-barrel shape with a hydrophilic lumen that allows charged ions and hydrophilic substrates to diffuse across the membrane passively. The main selectivity filter of solute diffusion is the size of the lumen pore, which is dictated by the extracellular loops and the type of amino acid distribution within the channel (18,19). For example, OmpF and OmpC trimers exhibit some preference towards cations and trimeric PhoE prefers anions, yet they are still considered as general diffusion porins; high-resolution models determined by X-ray crystallography were able to explain these differences in selectivity, with PhoE having a single Gly131 to Lys substitution, compared with OmpF, at the entrance of the pore (20-22). Certain porins exhibit some degree of selectivity towards the solutes they transport. For example, LamB, a trimeric porin, is involved in the uptake of maltose and maltodextrin oligosaccharides. However, substrate translocation still occurs by facilitated diffusion without the need for external sources of energy (23,24).

The OM also harbours a number of receptors involved in the active uptake of sparse nutrients from the cellular environment. For example, iron, an essential micronutrient, although abundant in nature is mostly present in a chemically unavailable, oxidised form. Due to this low abundance of biologically active metal ions, bacteria have evolved an arsenal of siderophores, low M_w chelating agents that scavenge ferric ions from the environment. These siderophores are too large to cross the OM via passive diffusion through general porins and instead utilise TonB-dependent outer membrane receptors, e.g. FepA and FhuA, in order to enter the cell (25). These receptors exhibit very high affinities (nanomolar range) towards these siderophores which, once bound to the receptor, are actively taken up through the receptor lumen (25,26). Structurally, these receptors form a 22 stranded β-barrel with an additional N-terminal region termed the plug domain that sits inside the lumen of the barrel and acts as a gate (27). Additionally, the N-terminus possesses a specific motif called the TonB box that associates with a periplasm-spanning IM protein TonB (25). This association results in the active uptake of the siderophore since TonB transfers the energy derived from the proton motive force (pmf) to the siderophore receptor, possibly through induced changes in receptor conformation (27,28) .

1.1.2 *Outer membrane biogenesis in E. coli*

All OMPs are synthesised in the cytoplasm and therefore must be transported across the IM and become inserted into the OM in their fully folded, functional state. Translocation of the the newly synthesised polypeptide chain across the IM occurs by the Sec-translocon complex into the periplasmic space where it associates with periplasmic chaperones, such as Skp, SurA and DegP (29). These chaperones act to prevent OMP misfolding and aggregation in the periplasm and in turn deliver the unfolded OMPs to their assembly points. The main chaperone associated with catalysing OMP folding and insertion into the OM is BamA (previously YaeT) that exists as part of the hetero-oligomeric β-barrel assembly module (BAM) complex and will be discussed in the next section in more detail (30). The periplasmic space is devoid of any readily available sources of chemical energy; hence these chaperones function along free energy gradient

pathways, uncoupled from ATP hydrolysis and the pmf. One model of an energy gradient between the periplasm and OM was proposed recently and stems from the very high thermodynamic stabilities of folded OMPs incorporated into the OM (31), wherein the free energy of OMP folding could lead to an energy gradient across the periplasmic space which may drive the unfolded OMPs into lipid bilayers, with the help of periplasmic chaperones along the free energy funnel (31). Numerous folding studies *in vitro* have demonstrated that OMPs are fully capable of proper folding and insertion into lipidic bilayers spontaneously without a source of mechanical energy or the action of chaperones. However, the rate of folding was deemed too slow to be physiologically relevant.

Accumulation of misfolded OMPs in the periplasmic space leads to the activation of the envelope-stress response, orchestrated by the σ^E – dependent stress response, initiated by membrane-associated protease cascades involving DegS and YaeL (32,33). Once activated, the cell envelope response dampens the expression of most OMPs and activates production of factors involved in the maintenance of periplasmic and OM integrity (34). The main chaperones under the control of the σ^E-dependent stress response are SurA, Skp and DegP (35,36).

Skp (seventeen kilodalton protein) is an important periplasmic chaperone implicated in preventing aggregation of unfolded OMPs during their periplasmic transit as they emerge from the Sec translocon (37). Additionally, Skp has been shown to interact with LPS molecules, which was shown to increase the efficiency of OMP folding *in vitro*, as demonstrated for the model OMP OmpA (38). The crystal structure of Skp revealed a trimeric jellyfish-like molecule with substrate OMPs occupying a central cavity, engulfed by the protruding “tentacles” (Figure 1-2 A) (39). Extensive biophysical studies involving NMR spectroscopy have characterised Skp-substrate complexes and revealed that the unfolded OMP undergoes chain compaction within the trimer lumen yet remains essentially unfolded upon association with Skp, forming a dynamic complex (40). Genetic knock-outs of *skp* give reduced levels of OMPs in the OM, though the cells are still viable (41). Skp has been shown to function in parallel with another periplasmic chaperone, SurA. Like Skp, deletion of the *surA* gene does not affect bacterial viability but results in

the accumulation of unfolded OMP intermediates. However, deletion of both *skp* and *surA* produces a synthetic lethal phenotype, suggesting that the function of at least one of these proteins must be maintained and that their functional roles could be overlapping (29). Biochemical evidence shows that SurA acts as a facilitator of OMP folding through a chaperone-like activity. The crystal structure of SurA has been solved and reveals an asymmetric dumbbell-like multi-domain protein. SurA possesses two parvulin-like peptidyl/prolyl *cis-trans* isomerase-like (PPIase) domains, P1 and P2, as well as an N-terminal domain and a small C-terminal domain (42). Together, the N- and C- terminal domains along with P1 form the core of the molecule that possesses a large crevice that can harbour hydrophobic peptides (Figure 1-2 B). Interestingly, the P2 domain is located far from the core domain and has been shown to possess some PPIase activity *in vitro*, yet the importance of PPIase activity in the physiological role of SurA is unclear (43). Studies of the specificity mechanisms of SurA show that the protein recognises motifs rich in aromatic residues such as Aro-X-Aro, i.e. motifs commonly found in OMP β-strands (44). Interestingly, the PPIase domain P2 is not involved in peptide recognition; peptides have been shown to associate mainly with the core domain with low micromolar affinities, suggesting a reversible complex formation which does not involve its PPIase domains (Figure 1-2 C) (44,45).

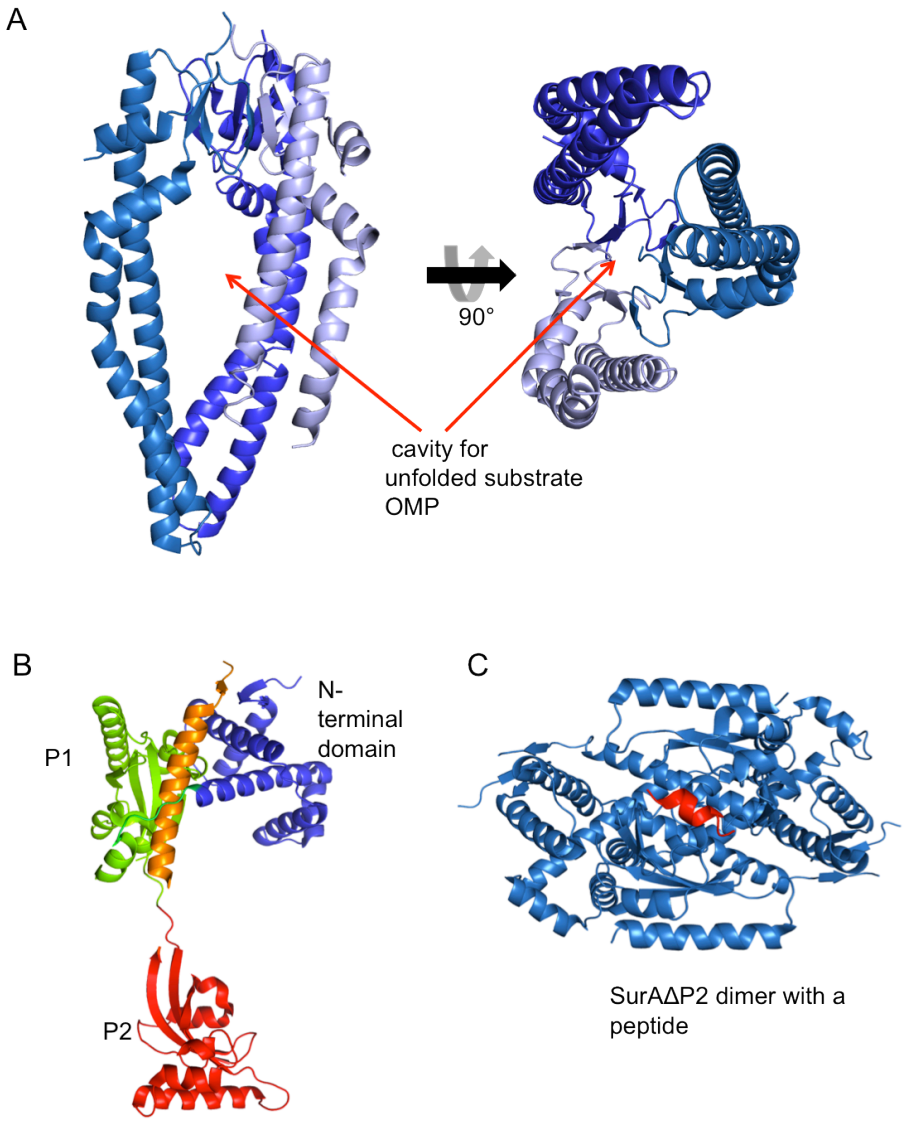


Figure 1-2 Insights into the function of periplasmic chaperones Skp and SurA. (A) Skp forms a trimeric structure with protruding α-helical extensions that form a central cavity. Unfolded substrate OMPs have been shown to fit inside this cavity as compact molten-globule-like molecules that lack native tertiary and secondary structure. Additionally, the lumen of Skp has been shown to associate with LPS molecules. (PDB ID: 1SG2) (B) High-resolution crystal structure of SurA reveals a highly modular protein consisting of 2 PPIase domains with P1 and the N-terminal domain forming the core of the molecule, and P2 (red) being located far from the core region. (PDB ID: 1M5Y) (C) Crystal structure of the core region of SurA harbouring a hydrophobic peptide (red). The peptide interaction is mediated by a SurA dimer. (PDB ID: 2PV3).

Another chaperone that is crucial to the maintenance of cell-envelope integrity is DegP, a multi-domain protein harbouring both chaperone and protease activity. Cells lacking the *degP* gene exhibit temperature-sensitive growth and a synthetic lethal phenotype when

surA is also non-functional (29). Functionally, DegP possesses two PDZ domains and a serine protease domain with a propensity to form higher-order oligomers: 6-mers, 12-mers and 24-mers. It is thought that oligomerisation of DegP is necessary for the regulation of its chaperone and protease activities. DegP has been shown to trap unfolded OMPs inside the cavity formed by the oligomers of the protein, with degradation of the substrate ensuing, depending on the level of misfolding in the trapped molecule (46,47).

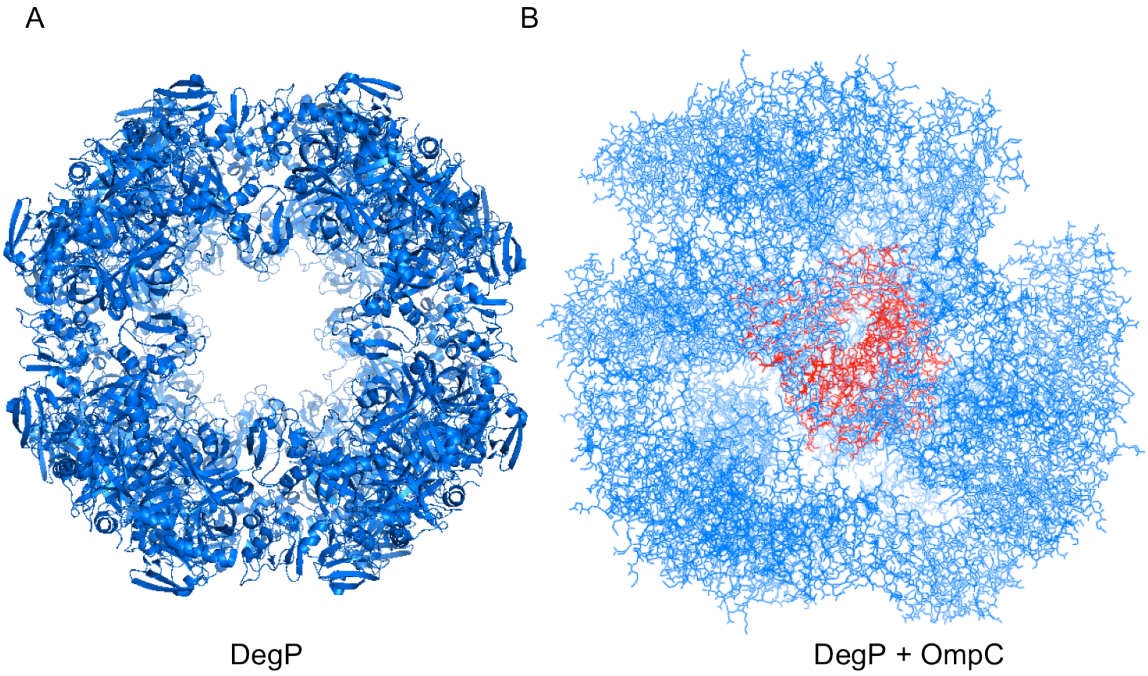


Figure 1-3 Structure of the DegP-24mer reveals a cage-like assembly that is able to trap misfolded OMP substrates for chaperone activity/degradation. (A) Crystal structure of DegP 24-mer reveals a cage-like oligomer capable of shielding misfolded substrates within its cavity (PDB ID: 3SC0) (B) Cryo-EM model of DegP-24mer (blue) harbouring the trapped OmpC (red) within the core of the oligomeric assembly (PDB ID: 4A8D).

1.1.3 *BAM complex*

The β-barrel assembly machinery (BAM) complex was initially identified in *Escherichia coli* as an OMP assembly machine after the role of the Omp85/YaeT protein in OM biogenesis was established (48,49,50). The BAM complex consists of five proteins, an integral OMP

BamA and four lipoproteins BamBCDE, that are anchored to the periplasmic face of the OM via BamA and/or covalent lipid anchor groups (30). Of the five proteins constituting the BAM complex, only BamA and BamD are essential for cell viability (49,51). Deletion of the other three members of the BAM complex leads to increased OM permeability and OMP folding and assembly defects (52,53). Further insights into the complex assembly reveal that the BAM complex can be separated into subcomplexes consisting of BamAB and BamACDE which have been shown to recombine into a single macromolecular complex with a 1:1:1:1:1/2 stoichiometry (54). *In vitro* reconstitution of the BAM complex from individually purified components in proteoliposomes (BamABCDE) was shown to drive OmpT folding and membrane insertion in an energy-independent manner, which suggests that the mechanism of OMP folding and membrane insertion is primarily encoded by the structure and composition of the BAM complex (55). However, a complete understanding of the mechanism of OMP biogenesis by the BAM complex requires further functional details about the roles of individual members of the complex and how the complex is assembled structurally. It is plausible that the BAM complex could form multimers *in vivo* thereby enhancing its functional capacity since there is some experimental evidence for oligomerisation of BamA (56).

1.1.3.1 BamA

BamA is itself an integral OMP with a C-terminal β-barrel and a soluble periplasmic region consisting of five POTRA (polypeptide transport associated) domains. BamA is a member of the highly conserved Omp85 protein family found in chloroplasts, mitochondria and bacteria. BamA is the only protein in the complex to span the entire OM whereas BamBCDE are anchored at the membrane via their acyl chains as well as interactions with BamA. Previous studies reported that several autotransporters show secretion deficiencies in BamA depleted cells, and numerous studies have shown BamA to be important in β-barrel protein assembly and OM integration (30,57). BamA, in addition to its integral β-barrel, possesses a periplasmic segment consisting of 5 tandem POTRA domains which have been implicated in OMP substrate recognition. The role of the POTRA domains in the function of BamA is only partially understood, but evidence suggests that they are involved in the protein-protein interactions required to form a

complete BAM complex, as well as substrate recognition and binding (30,58). Substrate OMPs are thought to be delivered to BamA by the periplasmic chaperone SurA that has been shown to associate with its periplasmic POTRA1 domain (59,60). Unfolded or partially folded OMPs are thought to present their C-terminal recognition sequences to the POTRA domains for their recruitment to the BAM complex (56). *In vitro* electrophysiology studies have shown that substrate binding to full-length BamA increases its channel activity through a conformational change relayed from the POTRA domains in the periplasm to the membrane inserted β-barrel (56).

Recent crystal structures of BamA from several bacterial species have revealed a wealth of information about the functional roles of this molecule in OMP biogenesis and folding (61,62). The difference in orientation of the BamA POTRA domains in the crystal structures suggests that their position with respect to the barrel depends on inter-molecular contacts with adjacent protein molecules. In one instance (61), the C-terminal POTRA5 domain occluded the periplasmic lumen of the BamA barrel, implying that these domains can modulate substrate access to the β-barrel cavity, which is essentially empty in all high-resolution models. One striking feature of BamA is the existence of a reduced hydrophobic region on one side of the barrel, namely between the first and the last β-strands (Figure 1-4 B highlighted in red). This weakly hydrophobic region is thought to introduce a local distortion into the lipid bilayer reducing its thickness, thereby allowing the emerging substrate OMPs, presumably in a partially folded state, to be successfully incorporated in the OM (61). This observation corroborates well with the biophysical experiments using denatured OMPs that exhibited increased folding rates with reduced lipid bilayer thickness *in vitro* (63). In addition to the reduced hydrophobicity at the barrel lateral gate, the crystal structure exhibited weakened inter-strand annealing between the first and last β-strands with only two hydrogen bonds seen in the crystal structure of *N. meningitidis* BamA (Figure 1-4B, PDB ID: 4K3B). This points to the possibility of lateral opening of the β-barrel, as demonstrated by molecular dynamics simulations (61). This lateral opening could serve as a release mechanism of OMP substrate polypeptides that have undergone initial folding steps within the BamA β-barrel cavity. Since the cavity inside the barrel is essentially empty, Albrecht *et al* have proposed that at least a single β-hairpin can be accommodated inside it (62).

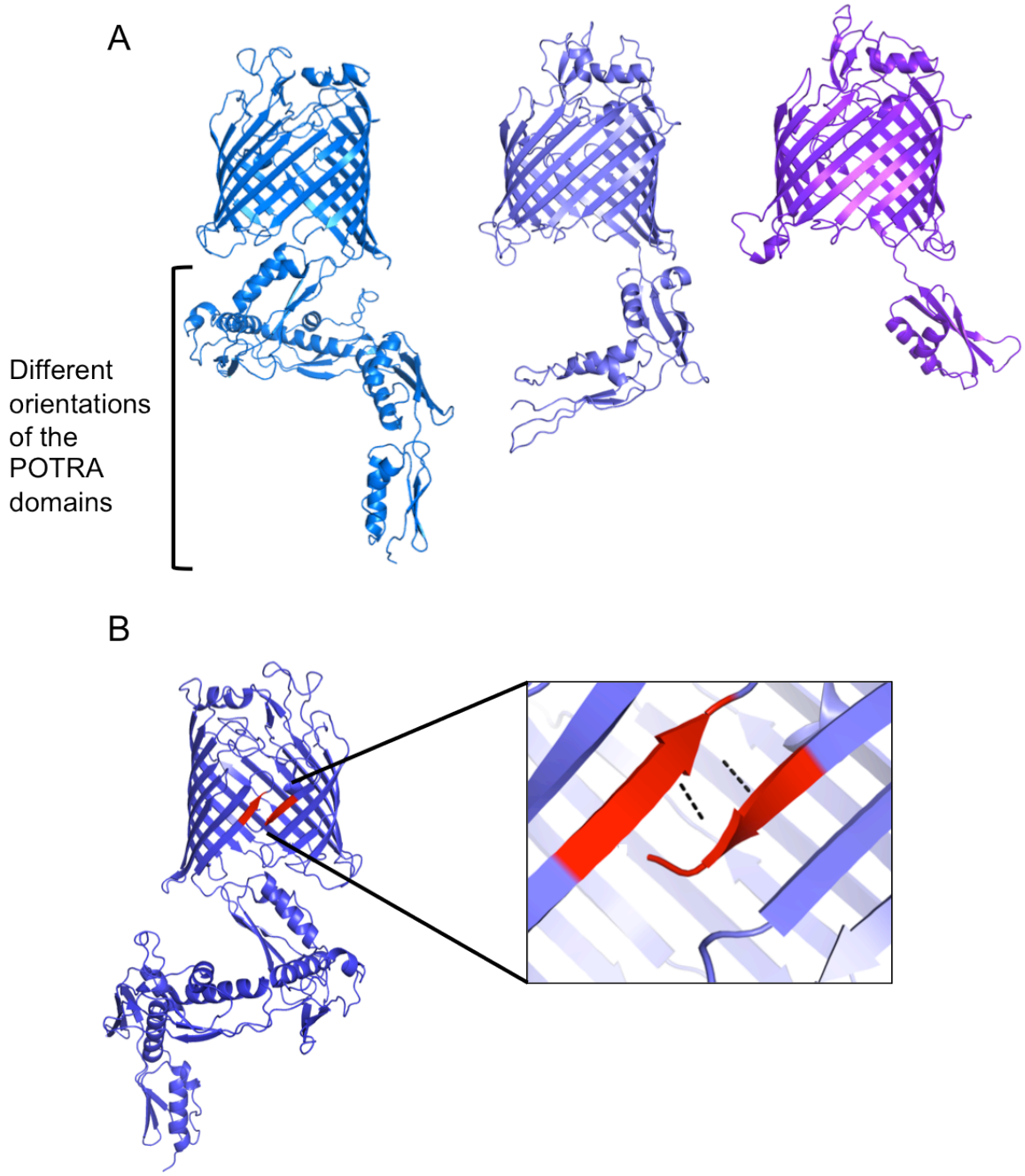


Figure 1-4 Crystal structures of BamA reveal the presence of a distorted β-barrel and highly mobile periplasmic POTRA domains. (A) Three high-resolution models of BamA showing relative orientations of POTRA domains observed *in crystallo*, suggest that these domains may also be involved in gating the barrel access from the periplasmic side (left to right PDB IDs: 4K3B, 4K3C, 4C4V). (B) Weak hydrogen bonding seen between β1 and β16 points towards putative lateral opening of the β-barrel in order to release or initiate folding of the substrate OMP (PDB ID: 4K3B).

1.1.3.2 BamB

BamB (previously YfgL) is a non-essential lipoprotein that forms direct contacts with BamA and has been functionally shown to increase the activity of the BAM complex *in vitro*, as well as interact with unfolded substrate OMPs *(*55,64*).* Cells lacking *bamB* exhibit reduced OMP levels and an elevated σ^E – dependent stress response, and together with the lipoprotein BamE, BamB is implicated in the folding and stability of BamA, as exemplified by the lethal *ΔbamB ΔbamE* phenotype (65). Moreover, double knockouts of *bamB* and *degP*, as well as *bamB* and *fkpA* also exhibit a synthetic lethal phenotype, which points to the importance of this protein in the context of OMP biogenesis (51). Several high-resolution crystal structures of BamB have been solved to date, which provide further functional insight into the role of BamB in OMP biogenesis (66,67). The protein possesses a conserved eight-bladed β-propeller fold, which is homologous to the WD40-repeat like domains that usually act as protein scaffolds modulating protein-protein interactions between multiple partners (Figure 1-5A) (68). Mutational and site-specific cross-linking studies have mapped a specific site in BamB involved in BamA-BamB interactions (69). Interestingly, the crystal structure of BamB solved by Heuck *et al* also suggests a putative mechanism of BamB-OMP substrate interactions through β-augmentation via the exposed β-strand of a propeller (70).

1.1.3.3 BamC

BamC (previously NlpB) is a lipoprotein that is highly conserved in γ-proteobacteria. However, phenotypic studies have found that *bamC* deletions have only mild effects on OMP levels and assembly, making it functionally elusive to characterisation (51). BamC is anchored to the BAM complex via its partner protein BamD and does not make direct contacts with BamA. The structure of BamC was first solved by Rosetta modelling and limited structural restraints derived from NMR spectroscopy (71). BamC possesses a modular structure made up of two domains with a mixed α/β fold as well as an unstructured 75 amino acid N-terminal domain. Each domain resembles a helix-grip fold and both folds are joined by a flexible linker with high inter-domain mobility (71). The structures of the N-terminal and C-terminal domains of BamC have also been solved by X-

ray crystallography to reveal a structure similar to that determined by NMR spectroscopy (72). Interestingly, structural studies show that the unstructured N-terminal region of BamC is the binding site for its partner protein BamD, which associates along the entire length of the molecule, as revealed by the crystal structure of the BamCD complex (73). Biochemical experiments have revealed increased proteolytic sensitivity of the BAM complex in *ΔbamC* cells, implying a role for BamC in the stabilisation of the BAM complex. Unexpectedly, it was also shown that the C-terminal domain of BamC could be exposed on the bacterial surface through an as-yet unknown mechanism (74).

1.1.3.4 BamD

BamD (previously YfiO) is an essential member of the BAM complex and, like BamB, directly interacts with BamA. Depletion of BamD lowers OMP levels in the OM and causes misfolding of BamA (53,55). Initial bioinformatic analyses of the BamD sequence suggested that the protein is composed of tetracorticopeptide repeats (75). The availability of a high-resolution crystal structure showed that BamD consists of five tetracorticopeptide repeats, each with a helix-turn-helix motif. BamD has been shown to associate with BamC as well as unfolded substrate OMPs (Figure 1-5 B) (55,73). Interestingly, a Δ*bamD Salmonella* strain is viable, but exhibits poor growth. This observation has been attributed to putative differences in the folding efficiency of the BamA β-barrel since a suppressor mutant $BamA_{F474L}$ exhibits reduced dependence on BamD depletion (65). BamD was shown to associate with BamA via POTRA5, the domain closest to the β-barrel, and it is therefore plausible that additional contacts are needed between BamD and the β-barrel for efficient functioning of BamA (30).

1.1.3.5 BamE

BamE (previously SmpA) is the smallest (10 kDa) and also the most conserved, yet non-essential protein across Gram-negative bacteria and has been functionally linked to stabilisation of the BAM complex (76). Deletion of *bamE* results in a compromised OM permeability and a consequent increased sensitivity to SDS, however the levels of OMP assembly in *ΔbamE* were only marginally reduced (76). High-resolution structures of

BamE have been obtained by both NMR spectroscopy and X-ray crystallography (Figure 1-5 C) (72,77,78). Interestingly, the fold of BamE is similar to that of β-lactamase inhibitor proteins, comprising an α/β sandwich. Knowles *et al* showed that BamE could associate with phosphatidylglycerol as well as BamD, which links BamE to the BAM complex (77). There is conflicting evidence about the oligomeric state of BamE, with experimental evidence for irreversible dimerisation and even hexamerisation (72,77). Kim *et al.* have suggested that the dimer is probably a kinetically trapped domain-swapped form of the protein; however Albrecht *et al.* propose that the dimerisation is physiological based on experiments in which full-length BamE was over-expressed and purified as a dimer from the OM fraction (72,78).

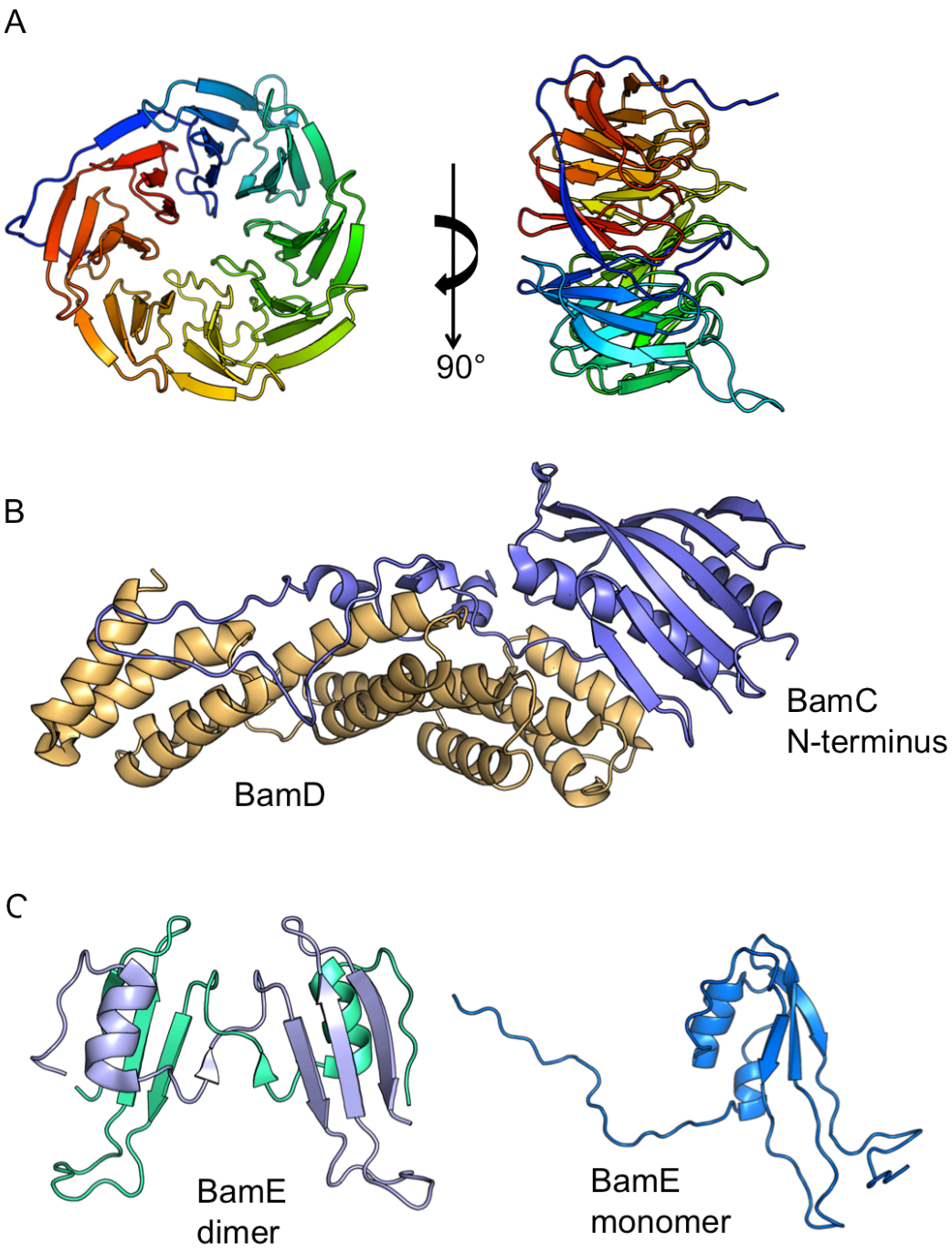


Figure 1-5 High resolution structures of components of the BAM complex. (A) Crystal structure of BamB showing the β-propeller fold as well as the presence of long, flexible loops between the propeller domains (PDB ID: 3PRW). (B) Crystal structure of the complex between BamD (golden) and the N-terminal region of BamC (purple). BamC possesses an extended, unstructured N-terminus which forms contacts over the entire BamD interface (PDB ID: 3TGO). (C) Dimer and monomer structures of BamE solved by X-ray crystallography (PDB ID: 2YH9) and NMR spectroscopy (PDB ID: 2KM7), respectively. The helix swap between the BamE monomers (teal and light blue) forms an irreversible dimer interface. The NMR structure has an extra α-helical motif not seen in the crystal structure.

1.1.4 *Omp85 family of proteins from eukaryotes*

Proteins belonging to the Omp85 family of proteins are found in both prokaryotic, as well as eukaryotic kingdoms, namely in mitochondria and chloroplasts. The presence of these proteins in eukaryotes emphasises strong evolutionary conservation of this class of proteins. Chloroplasts possess two paralogues of Omp85, namely Toc75 and Oep80 which are thought to have evolved from cyanobacterial proteins (79). Toc75 is primarily responsible for the import of preproteins into the chloroplasm from the cytoplasmic side of the organelle, yet the role of Oep80 is currently unknown. Both proteins have been shown to be essential yet functionally distinct (80). Mitochondria possess two major complexes in their OM. The first is the translocase of the mitochondrial OM (TOM) complex, made up of eight different proteins, and which is mainly responsible for the import of precursor proteins across the mitochondrial OM (81). Once inside the mitochondrial intermembrane space, the β-barrel OM proteins are recognised and assembled by the sorting and assembly module (SAM) complex (82). To date, the SAM complex consists of three main proteins, namely the integral membrane protein Sam50, and peripheral outer membrane proteins Sam35 and Sam37 (83,84). Sam50 was discovered as an essential membrane protein at the core of the SAM complex (85). Sam50 is a member of the Omp85 family of proteins as it was shown to possess a C-terminal 16-stranded β-barrel and an N-terminal POTRA domain akin to the Omp85 proteins. Interestingly, Sam50 is more closely related to the bacterial Omp85 proteins than to Toc75 (85). Sam35 was also shown to be essential for cell viability with a primary role of recognising substrate proteins and modulating the channel activity of Sam50 (86,87).

1.1.5 *Polypeptide transport-associated (POTRA) domains*

POTRA domains were identified as a conserved fold through extensive bioinformatic, sequence and secondary structure analyses where these domains are usually associated with transmembrane β-barrels in bacterial, chloroplast and mitochondrial proteins (88). Additionally, one POTRA domain was found to be present in the cell division protein FtsQ as well as its homologue DivIB; however, the role of this sole POTRA domain, as well as

the function of FtsQ/DivIB are still currently not understood (89). The residues most conserved within the POTRA fold are the hydrophobic amino acids comprising the core of the domain, as well as several Gly residues in loop regions, especially the GYF motif between the two α-helical segments. High degrees of sequence dissimilarity between these domains makes functional comparisons between them and their associated functions difficult.

POTRA domains are usually found at the N-terminus of the protein, and with the exception of the mitochondrial and chloroplast transport systems, are present in tandem repeats, with the number of domains varying between different translocation systems. The consequence of this variation in the number of POTRA domains is currently unclear. Omp85 proteins from *Myxococcus xanthus* possess up to 7 N-terminal POTRA domains, as does TamA from *Bdellovibrio bacteriovorus* (90,91)*;* Omp85 proteins from cyanobacteria*,* unlike their proteobacterial counterparts have 3 POTRA domains instead of 5 and the sorting and assembly module from mitochondria has only one POTRA domain (90,92,93). Functionally, POTRA domains have been implicated in substrate association and chaperone-like activity as well as formation of large membrane-associated macromolecular complexes (30,58,94).

Despite low sequence homology between individual POTRA domains, structural studies have shown that these domains have a very similar fold, consisting of two anti-parallel α-helices positioned on top of a three stranded β-sheet (30,95,96). Insights into the substrate association of POTRA domains come from the *B. pertussis* FHA/FhaC two-partner secretion system, where direct interactions between the N-terminal POTRA domain of FhaC and a structurally non-native polypeptide region of the substrate molecule, Fha30 (the N-terminal region of FHA) were demonstrated (94,97). This indicates that the POTRA domain of the FhaC transporter is able to associate with an unstructured translocation competent substrate, mimicking the translocation-like scenario *in vivo* (97)*.* These binding studies were later characterised in further detail using biophysical techniques and provided mechanistic insight into the function of these domains with β-augmentation proposed to be the primary mode of substrate association (94). Another study utilised NMR spectroscopy involving the POTRA12 domains from

BamA and hydrophobic peptides derived from the β-barrel domain of PhoE, a substrate of the BAM complex, to study POTRA-substrate interactions *in vitro* (58). Weak, transient interactions between PhoE-derived peptides and the BamA POTRA domains were detected by NMR spectroscopy, observing chemical shift changes caused by peptide binding and the resultant structural perturbation of amino acid residues in the POTRA12 domains (58). The mechanism of POTRA12-peptide association was consistent with the association of hydrophobic peptides along extended edges of the β-sheets and α-helices, which is in accord with the binding modes observed *in crystallo* during the crystallisation of POTRA1-4 (30). The high-resolution crystal structures of POTRA domains available to date provide a glimpse of the structural organisation of these domains, as well as the variety of topological orientations these domains assume, from bent to U-shaped and extended conformations (30,90,95,96). It is hypothesised that substrate binding is able to modulate the conformation of the POTRA domains, for example, in assisting β-hairpin formation during substrate translocation, however this effect has not been observed experimentally.

It has been suggested that polypeptide binding to the POTRA domains may induce β-strand formation in the substrate on the pretext that insertion of the substrate β-barrels is thought to occur via pre-formed β-hairpins. However, no experimental evidence exists to support this assertion. Functional studies of BamA showed that deletion of POTRA1 and POTRA2 does not drastically affect the growth rate of *E. coli* cells and some partial functionality of BAM is retained; however, POTRA3-5 are essential for cell viability (30). In contrast, BamA from *N. meningitides* can tolerate deletion of POTRA1-4 without any effect on cell growth, pointing towards species-specific differences in the importance of individual POTRA domains as well as species-specific differences in OM assembly mechanisms (98). Interestingly, deletion of the sole POTRA domain impaired but did not abolish the function of Sam50, and this POTRA domain has been implicated in substrate release from the SAM complex (99,100).

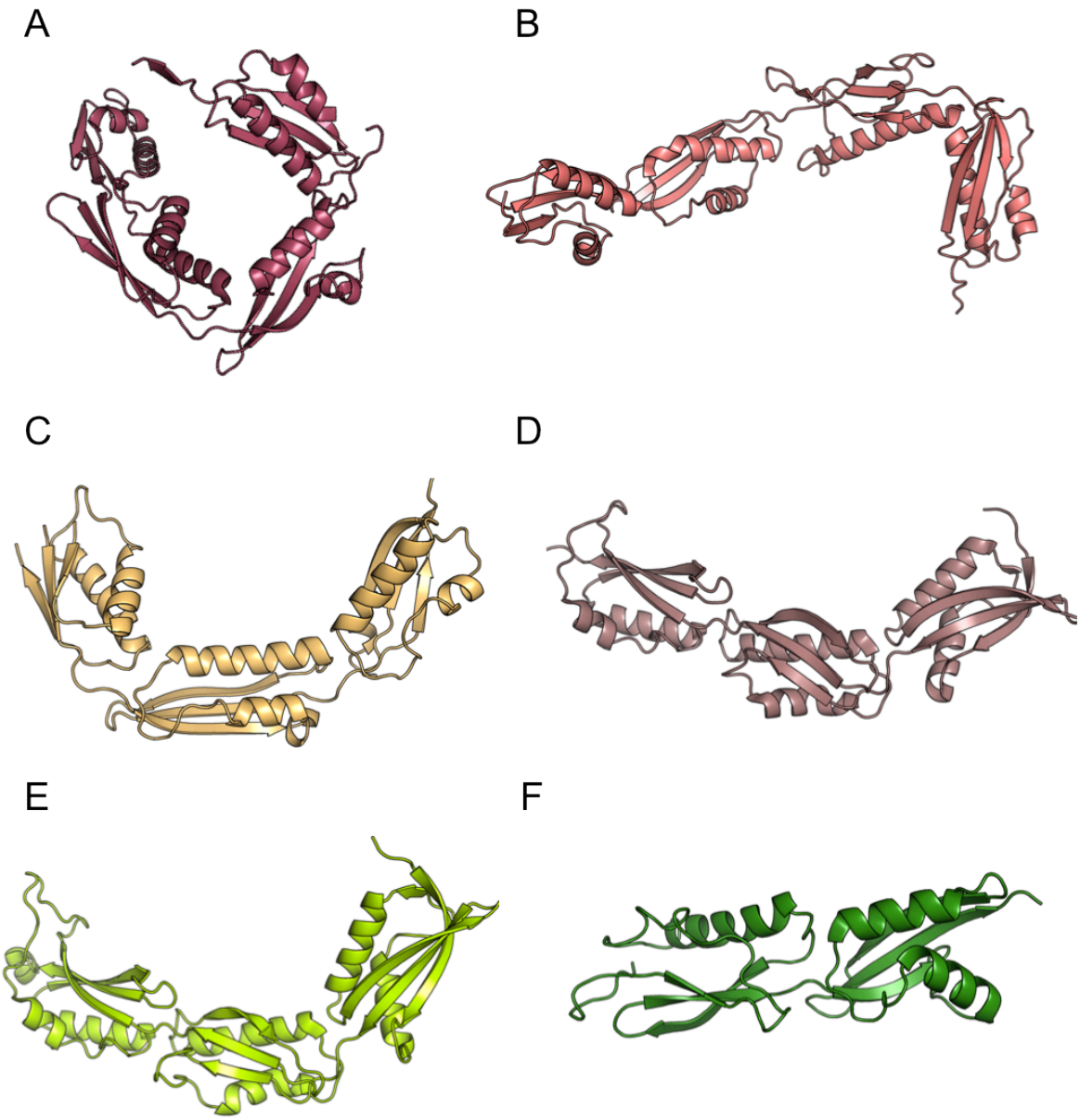


Figure 1-6 High-resolution structures of POTRA domains from Omp85 proteins of various bacterial species reveals diversity in the structural conformations of these domains. (A) BamA POTRA1-4 (PDB ID: 2QCZ) (B) BamA POTRA1-4 (PDB ID: 3EFC) (C) TamA POTRA domains. (PDB ID: 4BZA) (D) Omp85 POTRA from thermophilic cyanobacteria. (PDB ID: 2X8X) (E) POTRA domains from Anabaena Omp85 protein. (PDB ID 3MC8) (F) POTRA domains from FhaC. (PDB ID: 2QDZ).

1.2 Type V secretion systems

Bacterial proteins are synthesised in the cytoplasm and proteins that function on the cell surface or within the extracellular milieau must cross two lipid bilayers in order to reach their final destination. In order to efficiently and selectively translocate proteins across the bacterial membranes, bacteria have evolved numerous secretion systems. To date, seven families of secretion systems have been described in bacteria. One of these, the type V secretion system, is used widely by bacterial pathogens and functions to deliver virulence factors outside the cell (101).

1.2.1 *Two-partner secretion systems*

Two-partner secretion (TPS) systems are wide-spread throughout Gram-negative bacteria, and have been found in other kingdoms. TPS systems are frequently associated with bacterial virulence; however, they are mostly involved in the environmental adaptation of bacterial cells. The system consists of two proteins. One is the substrate molecule (TpsA) which is secreted outside of the cell and the other is the OM-embedded transporter protein (TpsB). The encoding genes almost always form an operon. Substrate molecules usually consist of a large, extended β-helix with a conserved N-terminal domain, called the TPS domain. This domain possesses the necessary determinants to initiate the secretion of the whole TpsA molecule across the OM. TpsB is usually an integral OM β-barrel that is involved in substrate recognition and association, at the same time providing a channel for the secretion process. Additionally, it is hypothesised that the β-barrel of TpsB initiates folding of the substrate through induction of local secondary structure formation as secretion occurs. The β-barrel transporter is a distant homologue of BamA-like Omp85 transporters. The widely studied TPS system FhaC/FHA is discussed in detail below.

1.2.1.1 FhaC-FHA

Filamentous haemagglutinin (FHA) is an adhesin from *Bordetella pertussis*, the causative agent of whooping cough, which is involved in host cell attachment and bacterial

colonisation (102). FHA is synthesised as a precursor molecule that becomes processed into its secreted form extracellularly by a subtilisin-like autotransporter SphB1 (103). The crystal structure of the N-terminal domain of FHA shows a typical β-helical fold which is capped at the N-terminus by three β-strands (104). Extensive mutational studies have identified the residues and motifs necessary for the secretion of FHA, which comprise conserved as well as non-conserved residues, relatively far apart in the protein sequence. This emphasises the complexity of the molecular recognition events that occur *in vivo* during substrate translocation, especially since only the non-native, unfolded form of the substrate can be recognised by FhaC (97). The structure of FhaC has been solved using X-ray crystallography and proved to be important in understanding the mechanistic details of the function of TPS systems in general (105). FhaC forms a 16-stranded β-barrel with two N-terminal POTRA domains which are involved in the interaction with the substrate (Figure 1-7A) (94). The barrel pore is occluded by an N-terminal helix as well as a large conserved loop, referred to as L6 (Figure 1-7B). Deletion of the conserved loop abolishes FHA secretion and reduces channel conductivity, implying that the conformational stability of the β-barrel may be compromised. Previous studies have concluded that L6 is accessible to protease cleavage from the extracellular side of the membrane in bacteria (106). Therefore, it has been proposed that the mechanism of FHA translocation by FhaC requires that upon substrate association the β-barrel undergoes conformational changes, such as L6 opening and N-terminal helix expulsion, in order to accommodate the translocating substrate.

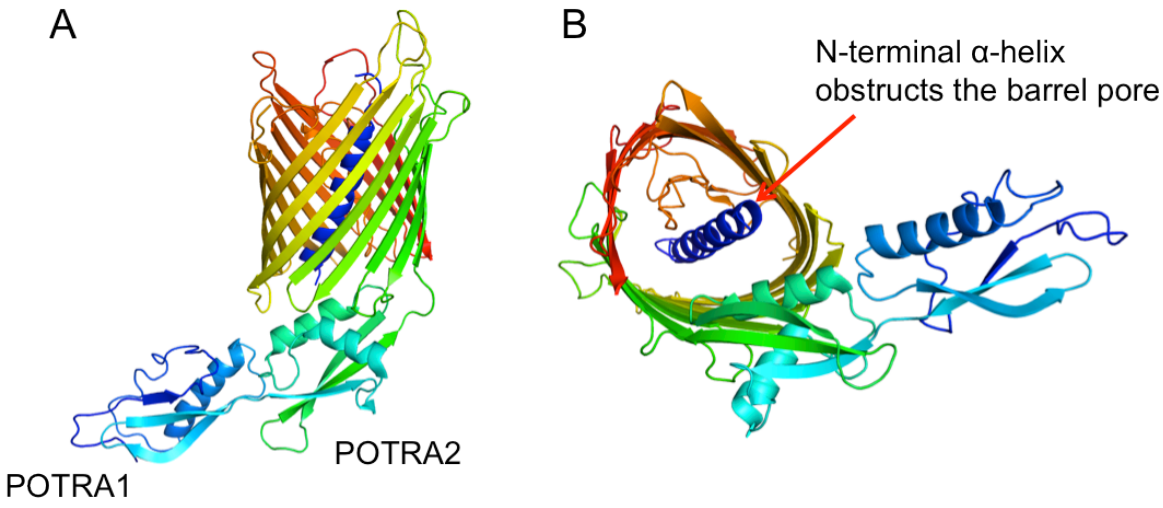


Figure 1-7 High-resolution crystal structure of FhaC shows an occluded β-barrel pore. The presence of the N-terminal helix within the barrel pore suggests that FHA translocation causes major structural re-arrangement within the molecule. (PDB ID: 2QDZ).

1.2.2 *Classical autotransporters*

Autotransporters (ATs) are OMPs that possess a C-terminal β-barrel domain and an N-terminal passenger domain consisting of a large β-helical structure and accessory domains harbouring a virulence function, e.g. protease activity or adhesion propensity. ATs are so-called because it was originally thought that the passenger domain was able to transport itself across the OM via its C-terminal β-barrel without the aid of any accessory proteins (107). However, it has long been suspected that additional proteins are involved in passenger domain translocation. ATs can be sub-divided further into two sub-classes, monomeric or trimeric, depending on their oligomeric state.

ATs possess three main structural features: an N-terminal signal peptide, followed by an extended passenger domain and a C-terminal β-domain which serves as an OM anchor for the whole molecule. Early observations showed that deletion of the β-domain abolished secretion of the passenger domain of an IgA protease AT. This finding led to the the term 'autotransporter' since it was believed that the molecule possessed all the

necessary components within itself for the translocation and secretion of the passenger domain across the OM. Most ATs are synthesised within the cytoplasm and are exported to the periplasmic space by the Sec system through the recognition of the N-terminal signal peptide by the signal recognition particle (SRP) targeting pathway (108). A subset of AT proteins (approximately 10%) possess an unusually long signal peptide (SP) (54 amino acids) in contrast to the common signal peptides found in secreted proteins (20-30 amino acids) (109,110). It has been demonstrated that long signal peptides stall the translocated ATs at the IM through interactions with the insertase YidC, thereby preventing misfolding of these large proteins in the periplasmic space (111). Truncation of the extended SP gives rise to misfolding of the EspP passenger domain *in vivo*, however β-domain folding and OM incorporation is unaffected (109). Additionally, it was proposed that ATs that possess extended SPs are unable to interact with SRP and thus do not undergo co-translational transport across the IM. Instead, their secretion occurs post-translationally and could involve an as-yet unidentified cytoplasmic chaperone (110).

1.2.2.1 The role of periplasmic chaperones in AT secretion

Once in the periplasmic space, AT SPs are cleaved by the leader peptidase, releasing the protein from the Sec translocon. Common periplasmic chaperones such as SurA, Skp and DegP have all been shown to participate in the periplasmic life-cycle of ATs and assist their translocation across the OM, as exemplified by studies with IcsA, an AT from *S. flexneri* (112). These chaperones are thought to target specific regions of the AT, however the molecular detail of their function is currently unknown. For example, Skp has been identified as the chaperone acting on full-length IcsA, although deletion of *skp* has only marginal effects on the levels of secreted passenger domain *in vivo*. The role of Skp was proposed to be the maintenance of passenger domain conformation in a partially unfolded or translocation-competent state for passage across the OM (113). In addition, photo-crosslinking studies have shown that Skp also associates with EspP in the early stages of AT translocation across the OM. Interestingly, Skp was shown to interact with the passenger domain as well as the β-domain during secretion of EspP (114). SurA, on the other hand, has only been shown to associate with the natively unfolded passenger domain of EspP *in vitro* and during OM translocation (114,115). Evidence suggests that

SurA acts at a later stage of AT biogenesis than Skp (114). Interestingly, secretion of Hbp exhibits little dependence on Skp and DegP yet is heavily dependent on SurA (116). These findings suggest that some redundancy is present amongst the periplasmic chaperones, as previously shown, and some ATs may rely on some chaperones and not others depending on the presence of specific sequence motifs. Like other OM β-barrels, targeting of AT β-domains to the OM is mediated by the BAM complex. Both BamB and BamD have been shown to associate with the AT β-domain of EspP *in vivo* and BamA has been shown to associate with both the β-domain as well as the passenger domain (60,114).

1.2.2.2 The AT translocation domain is a structurally conserved β-barrel

The β-domain is evolutionarily conserved in all ATs and is essential for passenger domain secretion (107). Several high-resolution crystal structures of AT translocation domains are available to date and provide invaluable mechanistic insights into the translocation mechanisms of these molecules. Among the first structures to be solved was the β-domain of NalP, an AT from *N. meningitides*. The high-resolution crystal structure of the β-domain of NalP revealed a 12-stranded β-barrel approximately 1 nm in diameter and a hydrophilic core. Additionally, the crystal structure showed that the barrel pore is obstructed by an α-helical segment immediately upstream of the β-domain (117). Electrophysiological measurements of NalP with the deleted α-helical linker exhibited higher conductance rates, and MD simulations of the translocation domain of NalP suggested that the helix serves as a regulator of channel activity and serves the role of a plug (117,118). The emergence of high-resolution models of the EspP β-domain as well as BrkA has revealed that the overall structure of the β-domain is highly conserved (119,120). Importantly, in all crystal structures the β-domain is present as a monomer ruling out multimer pore formation as a potential mechanism of AT translocation across the OM, as proposed by Veiga *et al* (121). The β-domain structures also reflect the different mechanisms of auto-catalytic cleavage of passenger domains from the β-domain, post-secretion. In the case of NalP, a member of the serine protease AT family (SPATE), passenger domain release occurs through an *in trans* cleavage using its serine protease domain (Figure 1-8 A,B) (122). Interestingly, other members of the SPATE family

undergo intra-barrel auto-cleavage releasing their passenger domain from the β-domain, as shown for studies with EspPβ (123). High-resolution crystal structures of EspP pre- and post-cleavage show that slight structural re-arrangements occur within the loop positions, with the catalysis mediated by conserved amino acids within the lumen of the barrel pore (Figure 1-8 C,D) (119,124).

A number of recent studies have identified motifs present in the β-domain that have been conserved to preserve the targeting of the AT to the OM (125). For example, one such motif comprises a glycine-aromatic residue pair on the two adjoining β-strands of the barrel that aid the folding of the domain through the facilitation of inter-strand contacts (126).

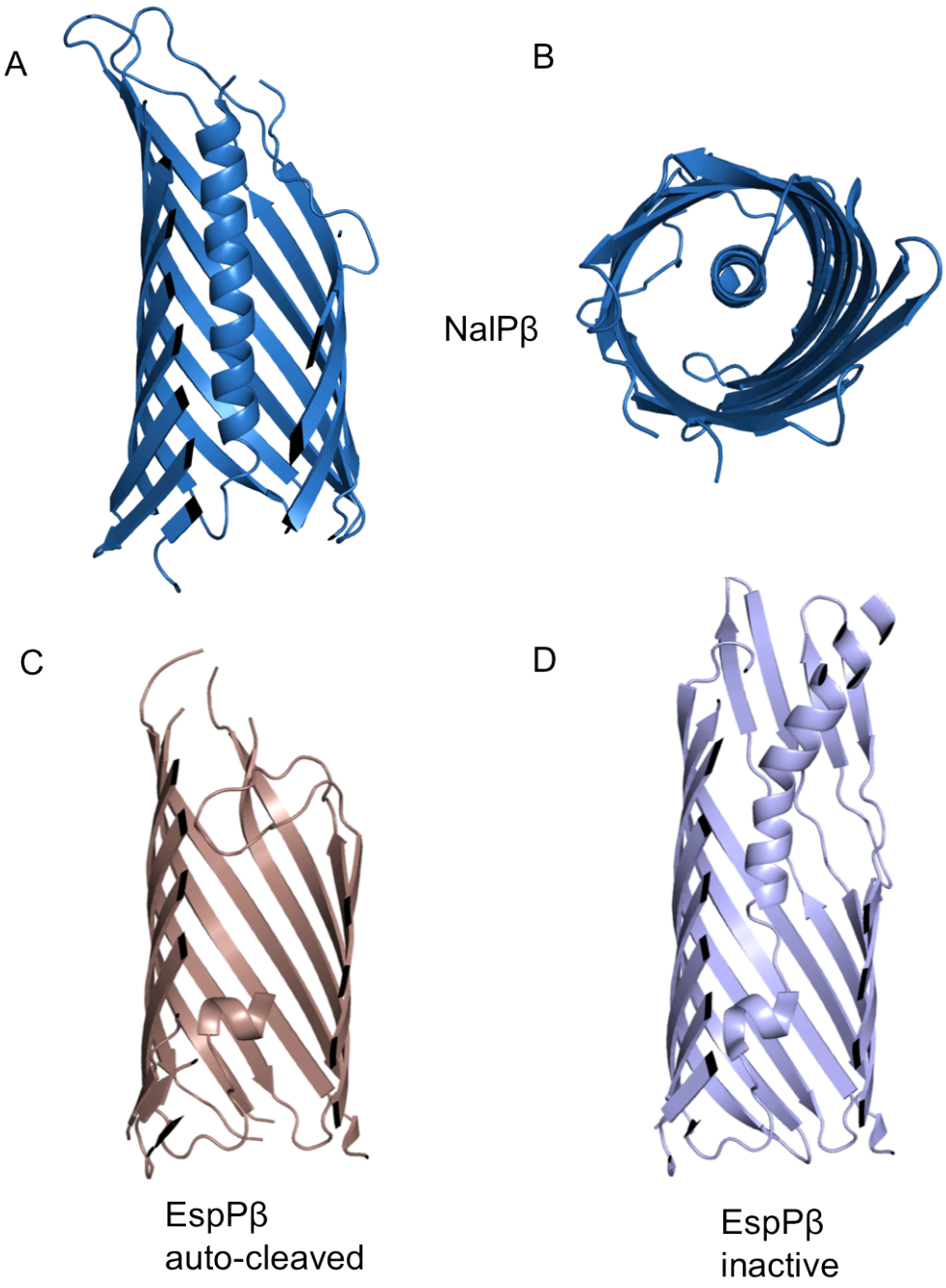


Figure 1-8 The crystal structures of two AT β-domains reveal different mechanisms of passenger domain release post-secretion. (A) and (B) High-resolution model of NalP β-domain showing the obstructed pore harbouring the full auto-chaperone α-helical peptide that has been cleaved at the bacterial surface (PDB ID: 1UYN) . (C) Crystal structure of EspP β-domain that has undergone intra-barrel autocatalytic cleavage. The barrel is gated by the remaining α-helical fragment as well as extracellular loops (PDB ID: 2QOM). (D) Crystal structure of an inactive EspP β-domain N1023A mutant that is unable to undergo autocatalysis (PDB ID: 3SLJ). This structure shows that, post-catalytic release of the passenger domain, some structural re-arrangement occurs within the extracellular loops in order to close the barrel lumen, compared with the model shown in (C).

The passenger domains of ATs form extended structures with a right-handed β-helical fold that extends from the cell surface. Bioinformatic analyses suggest that all AT passenger domains assume the β-helical fold (127) although these domain can also harbour additional accessory domains, e.g. a catalytic domain with protease activity in the case of EspP, Pet and IgA_1 protease or a lipase domain, for example EstA (128,129). These functional domains are located at the N-terminus of the passenger domain, whilst adhesin ATs usually possess simple β-helical structures with loops forming the functional sites on the molecule (Figure 1-9) (130).

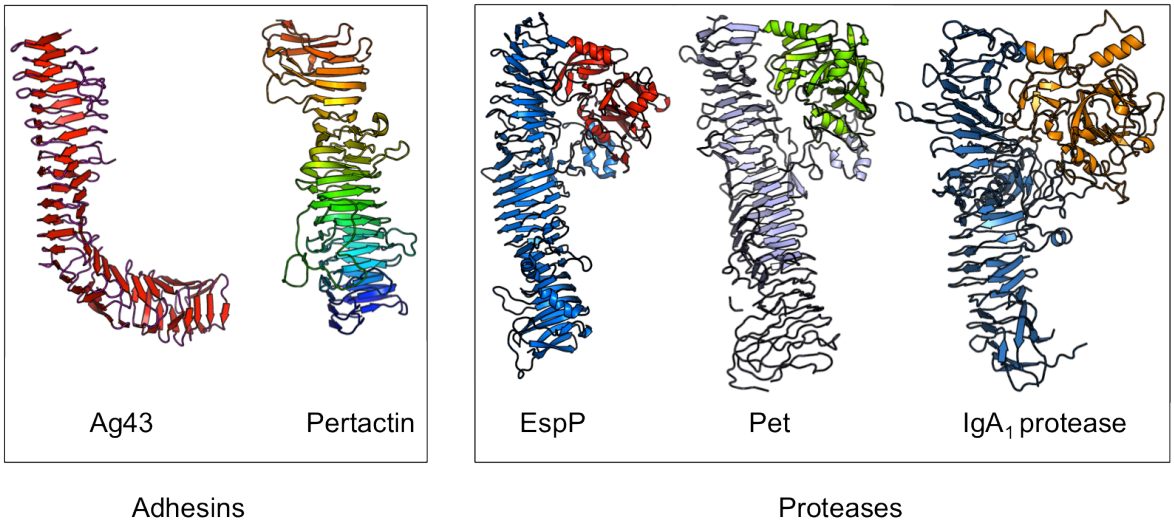


Figure 1-9 High-resolution crystal structures of AT passenger domain reveal a common right-handed β-helical architecture. Ag43 (PDB ID: 4KH3) and pertactin (PDB ID: 1DAB) both function as adhesin molecules assisting the formation of cell-cell contacts. Both structures reveal a simple β-helix with loops performing the functional adhesive properties by participating in (self) protein-protein interactions. ATs that belong to the serine protease family all possess an N-terminal proteolytic domain that belongs to a chymotrypsin protease family, as seen for EspP (red, PDB ID: 3SZE), Pet (green, PDB ID: 4OM9) and IgA_1 protease (orange, PDB ID: 3H09).

The conserved β-helical architecture of the passenger domains is thought to assist in domain secretion. The translocation of the passenger domain across the OM occurs vectorially in a C- to N-terminal fashion (131). Biochemical and biophysical studies have revealed that despite its structural uniformity, the passenger domain can be sub-divided into a stable C-terminal core region that exhibits fast folding kinetics *in vitro* and is

thought to initiate the folding of the rest of the passenger domain extracellularly (127,132) and an N-terminal region of the passenger domain that exhibits slow-folding kinetics and is unable to acquire its native structure in the absence of the C-terminal core region (133). This vectorial distribution of folding propensity across the domain is hypothesised to establish a folding gradient that would drive the translocation of the passenger domain polypeptide during secretion, or at least prevent back-sliding. Additionally, the slow folding kinetics over the whole passenger domain might prevent the formation of partially folded intermediates within the periplasmic space during AT biogenesis, that would stall the translocation process (133).

1.2.3 *Trimeric autotransporters*

Trimeric ATs are structurally distinct from the classical ATs. Functionally they act as surface adhesin molecules and are important for host colonisation and bacterial cell attachment. The process of oligomerisation is crucial for the adhesive activity of these ATs (134). It is thought that oligomerisation of trimeric ATs provides a multivalent site of attachment for target proteins. This enhanced adhesive property is beneficial when mechanical forces arising from the environmental niche are acting on the bacterial cell. Each individual AT is made up of an N-terminal signal peptide, followed by a passenger domain linked to an extended α-helical stalk region and a C-terminal translocation domain. The translocation domain of the trimeric AT is highly homologous to the translocation domain of the classical AT, inasmuch as it forms a 12-stranded β-barrel, with each monomer contributing four amphipathic β-strands and an α-helical coiled-coil stalk region, that are incorporated into the OM (Figure 1-10 A) (135). The frequency of the passenger domain-stalk region repeats is dependent on the individual AT. The passenger domain is thought to provide an additional oligomerisation site, leading to the formation of β-prism like structures (Figure 1-10 B).

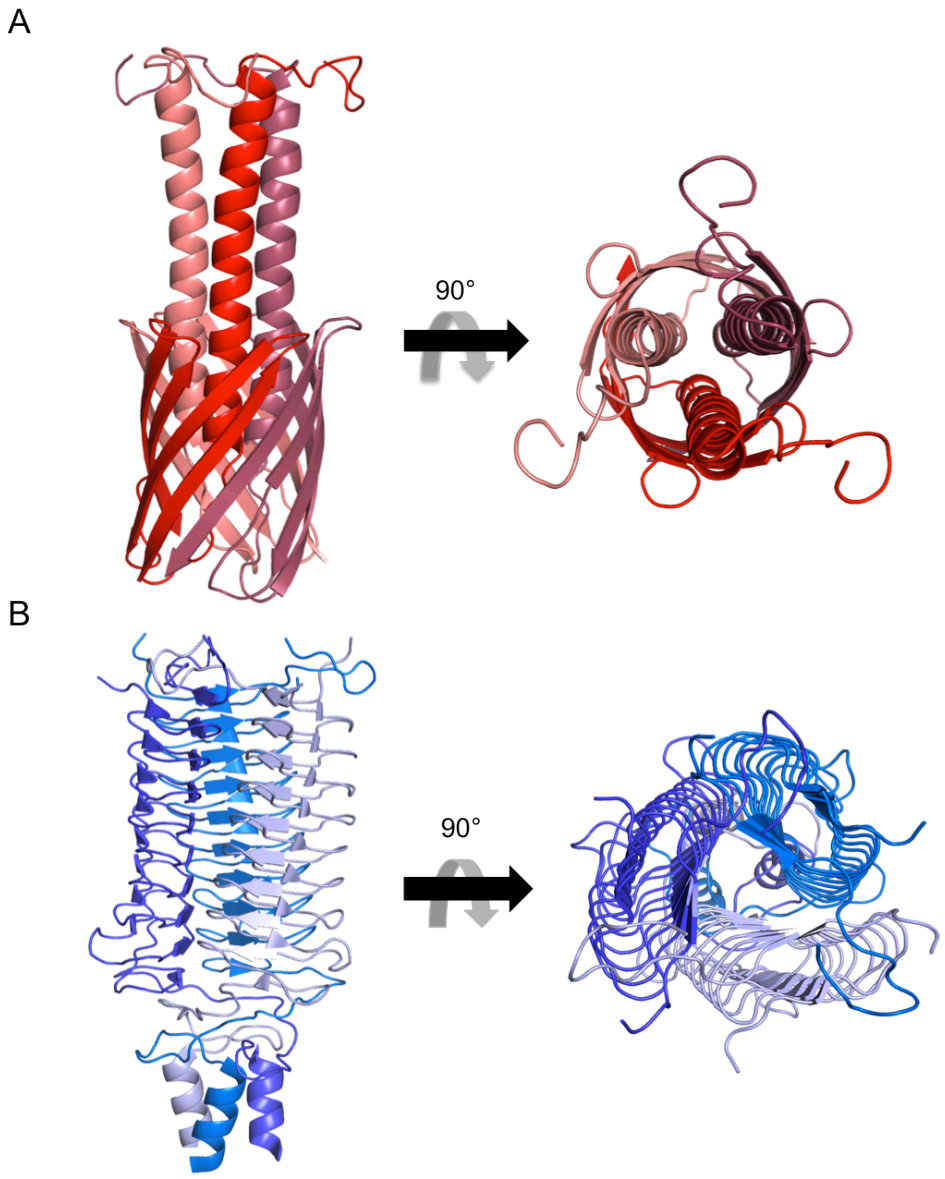


Figure 1-10 High-resolution models of YadA. (A) The structure of the trimeric translocation domain of YadA obtained by solid-state NMR spectroscopy. The β-barrel is formed by trimerisation of the C-terminal translocation domain, each monomer donating four amphipathic β-sheets (represented by shades of red). Additionally, each monomer contributes an extended α-helical motif leading to the formation of a coiled-coil within the barrel lumen that obstructs the channel (PDB ID: 2LME). (B) The head domain of YadA forms a trimeric β-helical assembly resembling a β-roll. The domains are stabilised by the formation of an α-helical coiled-coil motif that extends from the translocation domain shown in (A) as well as the trimeric β-helix (PDB ID: 1P9H).

1.2.4 *Current models of passenger domain translocation across the outer membrane*

The mechanism by which AT passenger domains are translocated across the OM has not been fully resolved, although several models have been proposed. The first model assumes that the passenger domain is threaded across the OM through the lumen of the β-barrel after it has been incorporated into the OM (Figure 1-11 A). However, this model is unlikely to be correct since the crystal structures of the β-domains of several ATs show that the lumen of the fully-folded β-barrel is too narrow to accommodate a polypeptide chain unless it is completely unstructured or possesses an α-helical conformation. This finding is in conflict with numerous studies which have suggested that passenger domain translocation can accommodate some degree of passenger domain folding during secretion (131,136).

A second model has been proposed based on structural studies of IgA protease β-domain purified from native membrane fractions. This domain was shown to form oligomers in solution and electron microscopy revealed the existence of ring-like structures with a central pore approximately 2 nm in diameter (137). This finding suggested that the AT β-barrels can oligomerise in the OM and form a central translocation pore through which the passenger domain could thread (Figure 1-11 B). However, the crystal structures of AT β-domains suggest that these domains crystallise as monomers, without any evidence of oligomerisation *in crystallo*. Moreover, the presence of the central α-helix within the lumen of the β-barrel suggests that passenger domain secretion does occur through its own translocation domain, perhaps prior to the full folding of the β-barrel (117,119).

An additional model was proposed based on the discoveries that the BAM complex, and mainly BamA, are essential components for AT biogenesis and secretion. BamA has been shown to associate with both the β-domain as well as the passenger domains of ATs *in vivo* and in light of recent structural studies of the function of BamA a new model of AT secretion was proposed (114,116). The lateral opening of the BamA barrel as well as the lipid bilayer distortion phenomena suggest that during the integration of the β-domain into the OM bilayer an intermediate complex between the AT β-barrel and the BamA

barrel could form, creating a large pore capable of translocating semi-unfolded passenger domain polypeptides. This model is consistent with the C- to N-terminal threading of the passenger domain, which may be aided by the POTRA domains of BamA, and the presence of the α-helical plug within the lumen of the AT translocation domain after the passenger domain has been secreted (Figure 1-11 C).

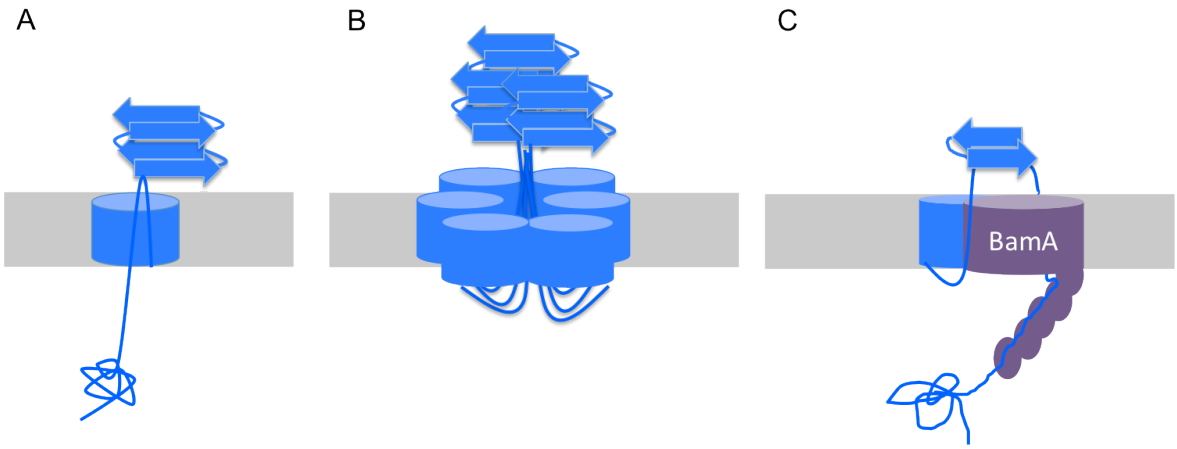


Figure 1-11 Models of AT passenger domain translocation across the OM. (A) The C-terminal β-domain is inserted into the OM and initiates the translocation of the passenger domain across the OM. (B) Several β-domains oligomerise in the OM forming a large central pore through which the passenger domain could be secreted outside the cell. (C) Passenger domain translocation occurs via an intermediate AT β-domain-BamA β-barrel with the assistance of the POTRA domains as well as the other members of the BAM complex (not shown).

1.3 The translocation and assembly module (TAM) complex

1.3.1 *TAM complex spans the periplasmic space*

Gram-negative bacteria possess an additional distant member of the Omp85 superfamily, named TamA, which possesses a C-terminal 22 stranded β-barrel and three N-terminal POTRA domains. *tamA* is present in an operon with another gene, termed *tamB* (138). Disruptions of *tamA* and *tamB* homologues in *P. mirabilis* lead to attenuated virulence in bladder and kidney infections in mice (139). Additionally, co-infection studies in mice with *C. rodentium* lacking TamA show that this protein is involved in pathogenesis since

the wild-type strain could out-compete the mutant strain in infection assays, and *tamA* was necessary for normal colonisation (138). Loss of *tamA* and *tamB* rendered *E. coli* cells sensitive to large hydrophobic antibiotics such as vancomycin and rifampicin, a phenotype usually associated with defective OM permeability caused by structural perturbations in the OM (138). However, the cell growth of the knock-out mutants was only marginally affected versus the wild-type strain. This, therefore, suggests that TamA and TamB are involved in some aspects of OM biogenesis or maintenance. TamA and TamB are anchored to the OM and IM, respectively (138). Protease shaving experiments suggest that TamB is at least partially located in the periplasm and the C-terminus of this protein possesses a conserved domain of unknown function, DUF490. The C-terminal region of DUF490 was implicated in the association with the POTRA domains of TamA (138). Co-immunoprecipitation experiments show that TamA and TamB form a complex that spans the cell envelope (Figure 1-12) (138).

1.3.2 *The TAM complex assists in the secretion of AT passenger domains across the OM*

Since Omp85 family proteins are involved in protein translocation it was hypothesised that TamA could be involved in some aspect of OMP translocation and membrane insertion. In fact, when *tamA* and *tamB* are deleted in *E. coli*, several defects in autotransporter secretion occur (138). Current experimental evidence implicates the TAM complex in assisting passenger domain translocation across the OM (138,140). Ag43, an adhesin involved in cell aggregation and surface attachment, was shown to accumulate in the periplasm in *ΔtamAB* cells in its full-length, unprocessed form and cells were unable to form a solid pellet post-centrifugation. Moreover, another AT, p1121, was also shown to be absent from the OM in cells lacking *tamA*.

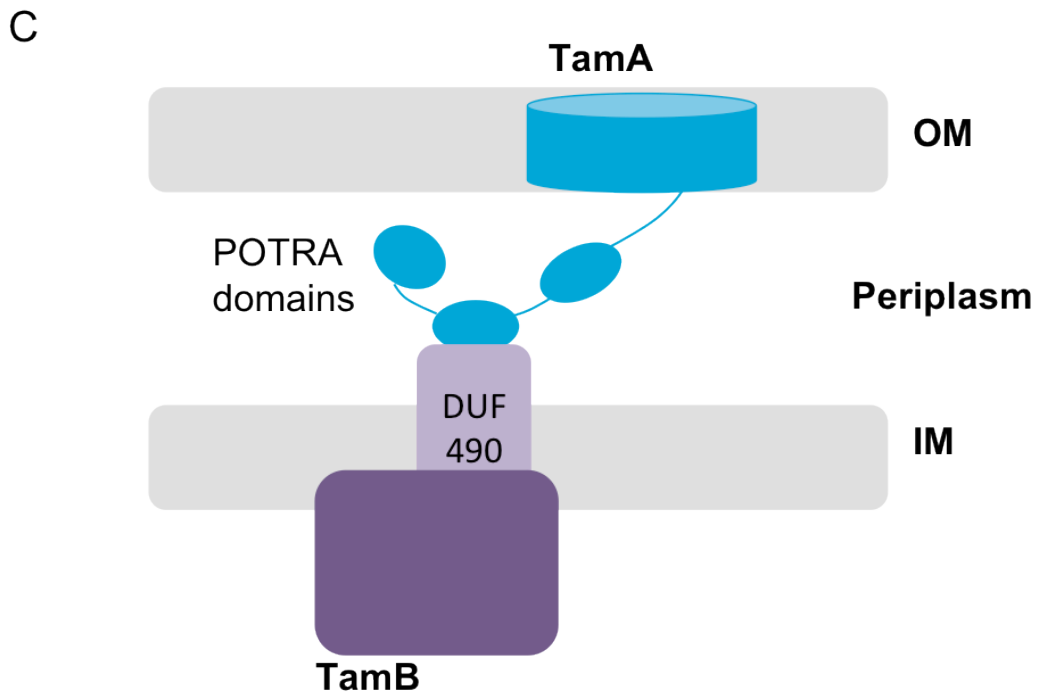


Figure 1-12 Current model of the cell envelope-spanning TAM complex. *tamA* and *tamB* are in the same operon and biochemical experiments have shown that both proteins form a complex that spans the periplasmic space. TamA is an integral OMP that belongs to the Omp85 superfamily of proteins, with three periplasmic POTRA domains. TamB is a large IM protein that has a conserved C-terminal domain of unknown function DUF490. TAM complex formation is mediated by the interaction between DUF490 of TamB and the POTRA domains of TamA.

1.3.3 *The high-resolution crystal structure of TamA reveals high structural homology to BamA*

The high-resolution crystal structure of full-length TamA has been reported (140). Despite low sequence identity between the β-barrel domains of TamA and BamA, several common features are apparent. The shape of the β-barrel closely resembles the barrel of BamA, rather than FhaC. The TamA barrel, similar to that of BamA, seems to exhibit a propensity for lateral opening, as indicated by fewer inter-strand backbone H-bonds between the first and last β-strands closing the barrel fold (Figure 1-13 A,B). In BamA, the possibility of lateral opening was hypothesised to be a mechanistic feature that would allow the release of folding substrate OMPs into the lipid bilayer. The functional

significance of lateral opening of the TamA β-barrel has not been previously addressed. In the case of BamA, blocking the lateral opening of the barrel using disulfide bonds between the first and last β-strands renders the molecule non-functional (141).

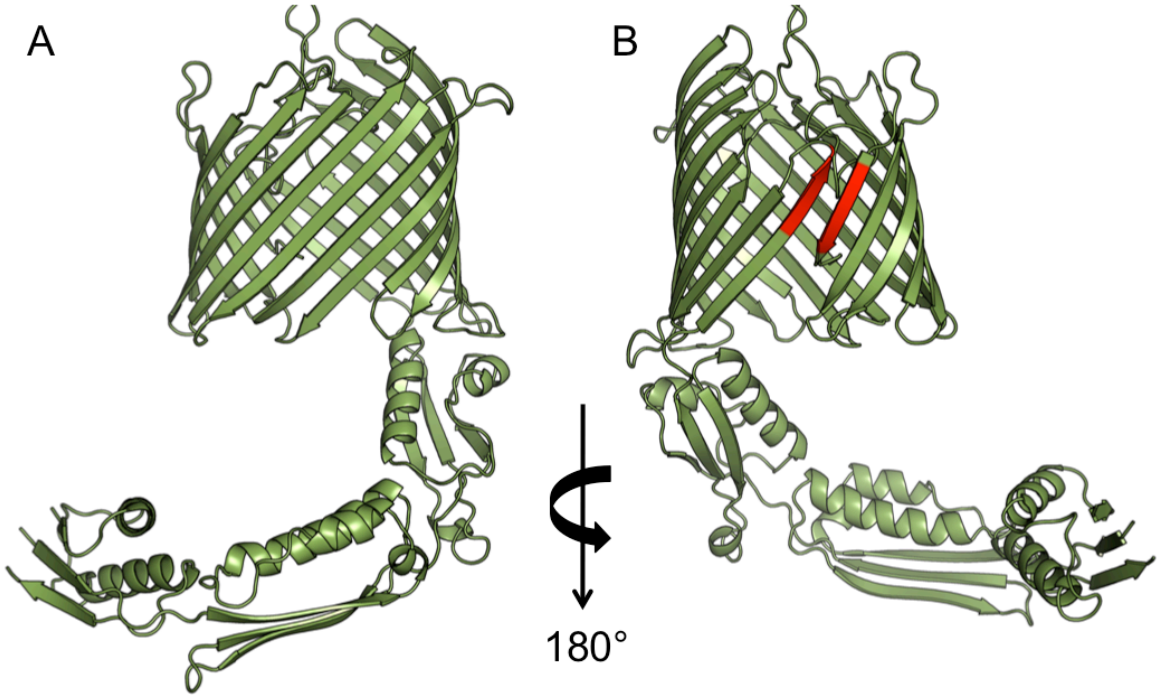


Figure 1-13 Crystal structure of TamA reveals a propensity for lateral opening on the β-barrel akin to BamA. The weak inter-strand hydrogen bonding is also present in the β-barrel of TamA, similar to the crystal structures of BamA. Additionally, unlike BamA, the relative orientation of POTRA domains does not indicate any POTRA-mediated gating propensity (PDB ID: 4C00).

Another structural feature that is apparent from the TamA crystal structure is the occluded β-barrel pore, shielded from the extracellular surface by a conserved loop L6, stretching between β-strands 11 and 12. The same loop occludes the BamA barrel from the surface side. This loop possesses a highly conserved motif VRGF/Y present amongst all members of the Omp85 family. Amino acid substitutions in loop L6 of BamA give reduced OMP protein levels in bacterial cells and this was experimentally linked to defects in the folding of BamA (142). Moreover, molecular dynamics simulations of BamA do not reveal

any obvious L6 mobility (61). Since L6 forms extensive contacts with the β-strands inside the barrel, it is thought that this loop serves a predominantly structural role.

Recently, it has been shown that TamA is a distant homologue of a newly discovered preprotein translocase, the archaic translocase of the *Trypanosoma brucei* mitochondrial outer membrane (ATOM) *(*143). Previously, no system had been identified for the protein import into the mitochondria of these parasites that lack the Tom40-like translocase found in most other eukaryotic cells. Through extensive bioinformatic analyses an Omp85-like protein was discovered in the trypanosomatid genome that was homologous to TamA (143). Biochemical experiments have shown that ATOM exists as a large macromolecular complex (700 kDa) in the mitochondrial outer membrane and is responsible for mitochondrial protein import (144). Little is known about the structure of ATOM, other than it contains a β-barrel domain. Sequence-based analyses did not identify any POTRA domains in ATOM (145).

1.3.4 *Aims*

The ultimate aims of this project were to obtain functional and structural insights into the role of the TAM complex in AT biogenesis. Previously demonstrated complex formation between TamA and TamB was to characterise quantitatively by attempting to delineate the interacting regions within the two proteins. This is important since little structural information is available regarding the participation of POTRA domains in macromolecular complex formation, for example in BAM complex formation. In addition, interactions of TamA and TamB with a known substrate AT, (Ag43), were explored to determine the mechanisms involved in substrate association and translocation and establish a model for the role of the TAM complex in passenger domain secretion.

2 Materials and Methods

2.1 Maintenance and growth of bacteria

All bacterial cells were grown in lysogenic broth (LB) media (recipe provided in Table 1). For protein overexpression LB media was supplemented with 4% glycerol and autoclaved. For NMR spectroscopy isotope labelling was carried out in M9 minimal media. All cells were grown at 37°C unless otherwise specified.

Table 1 Growth media

LB media	10 g/L tryptone
	5 g/L yeast extract
	0.17 M NaCl
M9 minimal media*	35 mM Na_2HPO_4
	22 mM KH_2PO_4
	8.5 mM NaCl
	18.7 mM NH_4Cl
	2 mM $MgSO_4$
	0.2 mM $CaCl_2$
	0.2 % D-glucose (anhydrous)

*for ^{15}N-isotope labelling, $^{15}NH_4Cl$ was used (CK Gas, UK)

For antibiotic selection ampicillin was used at a concentration of 100 μg ml^{-1} and kanamycin at 50 μg ml^{-1} (Melford, UK).

2.2 Preparation of chemically competent *E. coli* cells

Solution 1 (0.1 M $MgCl_2$) and solution 2 (0.05 M $CaCl_2$) were prepared and autoclaved. Prior to the start of the experiment, these solutions were placed on ice. *E. coli* cells (50 ml) were grown to an $OD_{600\ nm}$ between 0.4-0.6 and pelleted by centrifugation (4000 g) at 4°C. The pellets were washed twice with ice-cold solution 1 and centrifuged. The final pellet was resuspended with ice cold solution 2 supplemented with 15% (v/v) glycerol to

a final volume of 2 ml. Resuspended cells were aliquoted into 0.2 ml fractions, frozen in liquid nitrogen and stored in -80°C until further use.

2.3 Bacterial storage

All *E. coli* cells were stored in LB supplemented with 50% glycerol (v/v) at -80°C for prolonged storage. During experiments all over-expressing strains were kept on solid LB-agar plates at 4°C for approximately 6-8 weeks; after this period competent *E. coli* BL21 cells were re-transformed with appropriate expression plasmids.

Table 2 Bacterial strain table

Strain	Genotype and description	Source
Escherichia coli DH5α	F⁻ ϕ80*lac*ZΔM15 Δ(*lac*ZYA-*argF*) U169 *recA*1 *end*A1 *hsd*R17 (Γ_{k-}, m_k+) *pho*A *sup*E44 λ⁻ *thi*⁻1 *gyr*A96 *rel*A1	Invitrogen
Escherichia coli (DE3) BL21	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS* *λ DE3 = λ sBamHIo ΔEcoRI-B int: (lacI::PlacUV5::T7 gene1) i21 Δnin5*	New England Biolabs
Escherichia coli K-12 (BW25113)	*F´proA+B+ lacIq Δ(lacZ)M15 zzf::Tn10(TetR)/ fhuA2 glnV Δ(lac-proAB) thi-1 Δ(hsdS-mcrB)5*	Keio collection (146)

Table 3 Plasmids used in this thesis

Plasmid	Description	Reference
pET21a	Protein over-expression plasmid; resistance: Amp; promoter: T7	Invitrogen
pET28a	Protein over-expression plasmid; resistance: Amp; promoter T7	Invitrogen
pBad	Protein over-expression plamid; resistance: Amp; promoter: *araBAD*	Invitrogen
pPT123	TamA POTRA residues 22-265 in pET21a; His_6-C_t	This study
pPT12	TamA POTRA residues 22-190 in pET21a; His_6-C_t	This study
pPT1	TamA POTRA residues 22-102 in pET21a; His_6-C_t	This study
pDUF	TamB DUF490 domain, residues 926-1259 in pET28a; His_6-N_t	This study
pDUF1	TamB DUF490 domain, residues 926-1163 in pET28a; His_6-N_t	This study
pDUF2	TamB DUF490 domain, residues 963-1138 in pET21a; His_6-C_t	This study
$pTamB_{C80}$	TamB residues 1180-1259 in pET28a; His_6-N_t	This study
$pAg43_{N168}$	Ag43 passenger domain residues 55-220 in pET21a; His_6-C_t	This study
pSurA	resistance: Amp; promoter: T7; His_6-N_t	Dr Andrew Roe, Dr Katherine Beckham

2.4 Polymerase chain reaction (PCR)

2.4.1 *Amplification PCR*

PCR was used to amplify DNA fragments of interest. A typical amplification PCR reaction consisted of 1 μl Pfu Turbo DNA polymerase (Agilent, UK), 5 μl of 10x Pfu reaction buffer, 100 pmol of forward and reverse primers with appropriate restriction sites, 0.05 - 0.1 μg of template DNA and 0.2 mmol of each nucleotide from the dNTP mix (NEB, UK) and made up to a final volume of 50 μl. PCR reactions were carried out in the Eppendorf Mastercycler. The PCR amplification program consisted of an initial denaturation step at 95°C for 120 s, followed by 30 cycles of denaturation (95°C, 30 s), annealing (58°C, 30 s) and elongation (72°C, 180 s) with a final extension step (72°C, 600 s). PCR products were separated using agarose gel electrophoresis.

2.4.2 *Stratagene QuikChange Site-directed mutagenesis PCR*

PCR was also used for whole-plasmid site-directed mutagenesis. Mutagenesis reactions consisted of 1 μl of Pfu Turbo DNA polymerase, 5 μl 10X Pfu reaction buffer, 20 pmol of the forward and reverse primer, 0.5 μg-1 μg of template DNA plasmid and 0.2 mmol of each nucleotide. The PCR mutagenesis program consisted of 1 denaturation cycle (95°C, 30 s) followed by 16 cycles of denaturation (95°C, 30s), annealing (55°C, 60 s) and extension (68°C, 240 s). After the PCR reaction was complete, restriction enzyme DpnI (10 U) was added to the PCR products and incubated at 37°C for 4-5 hours. After the incubation, 5 μl of the PCR reaction was used for transformation into chemically competent DH5α cells.

2.4.3 *Oligonucleotide primers*

All oligonucleotide primers used in this thesis were synthesised by Eurofins Operon.

Table 4 Primer table

Primer	Additional feature	Sequence
$TamA_{POTRA}$ R (residues 22-265)	5' XhoI site	CGCAACTCGAG ACG TTC GCC ACT GTT ATA ATC AAT AT
$TamA_{POTRA12}$ R	5' XhoI site	CGTAACTCGAGTTC TGT TCG CGG CGA AAC C
$TamA_{POTRA1}$ F (mutagenesis)	Gly31Cys	CCGTCTACAGGTCGAGTGCTTATCGGGACAGCTGG
$TamA_{POTRA1}$ R Gly31Cys (mutagenesis)	Gly31Cys	CCAGCTGTCCCGATAAGCACTCGACCTGTAGACGG
$TamB_{C80}$ F (mutagenesis)	NdeI insert	GCGCAAAGTGGCCAGCATATGGGTAAAATCGGCGAGAC G
$TamB_{C80}$ R (mutagenesis)	NdeI insert	CGTCTCGCCGATTTTACCCATATGCTGGCCACTTTGCGC
$Tam_{DUF490\Delta20}$ F (mutagenesis)	Stop codon insert	GCCTGATGCCTAAGCTATAGCTGGAAGCCGTGTCTGG
$Tam_{DUF490\Delta20}$ R (mutagenesis)	Stop codon insert	CCAGACACGGCTTCCAGCTATAGCTTAGGCATCAGGC
$Tam_{DUF490\Delta35}$ F (mutagenesis)	Stop codon insert	CGGCGTGGGTATATTTGACTCTTAAGCAACACTCACGTT ACGTTATC
$Tam_{DUF490\Delta35}$ R (mutagenesis)	Stop codon insert	GATAACGTAACGTGAGTGTTGCTTTAGAGTCAAATATAC CCACGCCG
DUF_{trunc} F (mutagenesis)	Residues 926-1163. Stop codon insert	GCGACAGTGCGGCAATGTAGTCGATGCTGATTGGTTTG
DUF_{trunc} R (mutagenesis)	Residues 926-1163. Stop codon insert	CAAACCAATCAGCATCGACTACATTGCCGCACTGTCGC
$DUF_{963-1138}$ F	NdeI site	TCAGCATATGATGGATGTATCGCCAGAT
$DUF_{963-1138}$ R	XhoI site	GGTACTCGAGCGACATCGCCGGGTCAGA
$Ag43_{N168}$ F	NdeI site	GTCTCATATGGATATCGTCGTCCATCCGGG
$Ag43_{N168}$ R	XhoI site	GCTGCTCGAGTGAGCCATTTTTGTTAATCGTCGTG

2.5 Restriction enzyme digest

All restriction enzymes were purchased from New England Biolabs (NEB), UK. Typically, 10-20 μg of DNA was mixed with 1 μl of 10x Cutsmart buffer (NEB, UK) and appropriate restriction enzymes (20 U), made up to 15 μl and incubated at 37°C for 3-4 hours. Digested DNA was then electrophoretically separated on an agarose gel. In the case of plasmid DNA, 5 U of calf intestinal alkaline phosphatase was added after 2 hours of digestion time.

2.6 Agarose gel electrophoresis

Agarose gels were made using 0.8% (w/v) agarose and TBE buffer (0.17 M Tris-HCl, 0.2 M borate, 5 mM EDTA, pH 8) followed by microwaving until the agarose was dissolved. The mixture was allowed to cool to approximately 45-50°C before GelRed (Cambridge Bioscience, UK) was added (1:10,000) and poured into the gel tray. Once the gel was set, TBE buffer was added to cover the gel entirely. All DNA samples were made up with 5x DNA loading dye (Bioline, UK). 1 kb+ DNA ladder (Invitrogen) was loaded as a molecular size reference. Gels were run at 60 V for 40-60 min. Gels were visualised using a UVIpro Gold transilluminator (UVItec, UK).

2.7 Agarose gel DNA extraction

DNA bands of interest were excised from agarose gels under a transilluminator and extracted using a Qiagen gel extraction kit (Qiagen, UK) following the manufacturer's instructions. DNA was eluted in 30-50 μl of elution buffer (10 mM Tris-HCl, pH 8.0) and stored at -20°C.

2.8 DNA fragment ligation

0.1-1 μg of digested plasmid DNA was mixed with 3-5 fold excess of digested ligation fragment, 1.5 μl of 10x T4 DNA ligase buffer (NEB, UK) and 1 μl of T4 DNA ligase (5 U). The

ligation reaction volume was made up to 15 μl and incubated over night at 16°C. 5 μl of the reaction was used for the transformation into chemically competent DH5α cells.

2.9 Transformation of chemically competent *E. coli* cells

Competent *E. coli* cells were thawed rapidly at room temperature and 0.5 -1 μl of pure plasmid (50-80 μg/ml) was added to 50-100 μl of cells. In the case of a ligation reaction, 5 μl of the ligation mixture was added to the cells. Cells with the added plasmid DNA were stored on ice for 30 min, followed by a heat-shock at 42°C for 45 s. Following heat-shock, the mixture was placed on ice for 3-4 min and 0.5-0.7 ml of LB media was added to the cells. Cells were then placed at 37°C for 45-60 min for recovery and plated out on agar plates with appropriate selection antibiotics; transformed cells were grown overnight at 37°C. In the case of ligation reactions, recovered cells were gently pelleted by centrifugation (500-800 g), cell suspension was spread onto an LB-agar supplemented with appropriate selection antibiotics.

2.10 SDS-PAGE

Samples were mixed with Laemmli buffer (200 mM Tris-HCl pH 6.8, 8% (v/v) SDS, 0.4% (v/v) bromophenol blue, 40% (v/v) glycerol, 4.7% (v/v) β-mercaptoethanol) in a 1:3 ratio, boiled at 95°C for 5-10 minutes and centrifuged (5000 g) for several minutes prior to loading on the gel. Gels were run in Tris-glycine running buffer (0.02 M Tris-HCl, 0.25 M glycine, 5% (w/v) SDS) at 30 mA for 45-60 minutes, stained using Coomassie stain (40% (v/v) methanol, 10% (v/v) acetic acid, 0.5 g Coomassie blue R250) and subsequently destained with destain solution (40% (v/v) methanol, 10% (v/v) acetic acid). A broad range molecular weight marker (2-212 kDa, NEB, UK) was run alongside the samples on the gel.

2.11 Western blotting

Following SDS-PAGE, proteins were transferred from the gel onto nitrocellulose membranes using an ECL semi-dry transfer unit at 30 mA for 1 hour. After the transfer, the membrane was blocked in blocking buffer consisting of 20 mM Tris-HCl, 200 mM NaCl, 0.01% (v/v) Tween 20, 5% (w/v) skimmed milk (Marvel, UK) for 2 hours at room temperature. Primary anti-Ag43 antibody (1:2000) was applied to the nitrocellulose membrane for 1 hour at room temperature and agitated; excess unbound antibody was washed using 20 mM Tris-HCl, 200 mM NaCl, pH 7.2, 0.01% (v/v) Tween 20 several times. The membrane was then incubated with a goat anti-rabbit secondary antibody (1:10,000) coupled with horseradish peroxidase (HRP) for 1 hour whilst shaking at room temperature and washed as described above. The blot was developed using enzyme-linked chemiluminescence (ECL) SuperSignal West Picao chemiluminescent substrate (Thermo Scientific, UK).

2.12 Protein over-expression and purification

The following describes the general protocol for over-expressing protein construct used in this thesis. Details pertaining to individual constructs are described in the appropriate Results chapters. Chemically competent *E. coli* (DE3) BL21 cells were transformed with appropriate plasmids for protein over-expression. Overnight cultures were added to over-expression medium at a 1:100 ratio. Cells were grown to an $OD_{600\ nm}$ of 0.4-0.6 at 37°C and over-expression was induced by adding isopropyl β-D-1–thiogalactopyranoside (IPTG). The cell pellet was collected by centrifugation at 4400 x g and re-suspended in buffer A (20 mM Tris-HCl, 10 mM imidazole, 0.5 M NaCl, 5% (v/v) glycerol, pH 7.5) supplemented with complete protease inhibitor cocktail (Roche) plus lysozyme (2mg ml^{-1}) and lysed by sonication. Cell debris was cleared by additional centrifugation at 46,000 x g and the supernatant was passed through a nickel-charged HisTrap HP column (GE Healthcare). The bound fractions were collected after elution with buffer B (20 mM Tris-HCl, 350 mM imidazole, 0.5 M NaCl, 5% (v/v) glycerol, pH 7.5). Fractions containing the

protein of interest were pooled and dialysed in appropriate buffers and purified further on a Superdex S200 or Superdex S75 gel-filtration column equilibrated in the same buffer.

2.13 Electrospray ionisation mass spectrometry (ESI MS)

ESI MS was performed on an Agilent Q-TOF (6520) mass spectrometer at the FingerPrints proteomics facility (Dundee, UK). Protein sample (1 mg ml^{-1}) was dialysed against 10 mM sodium phosphate, pH 7.5, followed by dialysis in dH_2O.

2.14 Calibration of the analytical SEC column

An analytical Superdex S200 GL 10/300 column (GE Healthcare) was calibrated using a set of molecular weight standards (Sigma, UK). Prior to equilibration, the column was equilibrated with 50 mM Tris-HCl, 200 mM NaCl, 0.015% (v/v) Triton X-100 with the flow rate set to 0.7 ml min^{-1}. Protein markers (12 kDa – 2 MDa) were purchased from Sigma (UK) and solutions made up in the calibrant buffer to a final concentration of 1 mg ml^{-1} were injected onto the column (500 μl injectant volume). The elution volume of each protein was plotted against the log M_w in order to obtain a calibration curve.

2.15 MTSSL-labelling of TamA POTRA domains

$TamA_{POTRA1(Gly31Cys)}$ was produced by site-directed mutagenesis and transformed into *E. coli* (DE3) BL21 strain. Proteins were over-expressed and purified to homogeneity using nickel-affinity and size-exclusion chromatography supplementing the buffers with 1 mM TCEP during the nickel-affinity stage. Purified proteins were dialysed into 50 mM sodium phosphate buffer, 100 mM NaCl, pH 7 and mixed with a 5-fold excess of the spin-label MTSSL ((1-Oxyl-2,2,5,5-tetramethylpyrroline-3-methyl) methanethiosulfonate, Enzo Life Sciences, UK). Proteins were incubated with the paramagnetic tag overnight at 4°C and passed through the SEC column to remove excess, unbound spin-label.

2.16 Nuclear magnetic resonance (NMR) spectroscopy

2.16.1 *Analysis of $TamA_{POTRA123}$, $TamA_{POTR12}$ and $TamA_{POTRA1}$ using NMR spectroscopy*

All NMR experiments were performed on a Bruker AVANCE 600 MHz spectrometer fitted with a ^{15}N, ^{13}C cryoprobe. Fast HSQC spectra were acquired from ^{15}N labeled $TamA_{POTRA12}$ and $TamA_{POTRA1}$ in 100 mM sodium phosphate, pH 6.9, 0.015% (v/v) Triton X-100, 10% (v/v) D_2O in the absence and presence of excess unlabelled $TamB_{DUF490}$ (stoichiometric ratio roughly 1:4) to visualise $TamA_{POTRA12}$-$TamB_{DUF490}$ interactions. Spectra were processed with TopSpin and analysed with CCPNmr analysis software (147). The Fast-HSQC spectrum of ^{15}N-labelled $TamA_{POTRA123}$ was acquired with a 90 μM sample in 75 mM sodium phosphate, pH 7, 100 mM NaCl, 0.5-1% (v/v) glycerol at 295 K.

Double-resonance backbone assignment was carried out for ^{15}N-labelled $TamA_{POTRA1}$ in 50 mM sodium phosphate, 10% (v/v) D_2O pH 6.9 (protein concentration ranging between 0.75-1.2 mM). 3D ^{1}H-^{15}N HSQC-NOESY (mixing time 100 ms) and 3D ^{1}H-^{15}N HSQC-TOCSY (mixing time 60 ms) spectra (148) were acquired and processed with AZARA (W. Boucher, www.bio.cam.ac.uk/azara) using the MaxEnt method, and assignment was carried out with the CCPNmr software package to 82% completeness.

2.16.2 *Analysis of $Ag43_{N168}$ using NMR spectroscopy*

For ^{15}N-labelled $Ag43_{N168}$, fast HSQC spectra were acquired at 278 K with 100 μM protein in the absence and presence of unlabelled 1 mM $TamB_{C80}$ and 2 mM $TamA_{POTRA1}$. Binding studies between $TamA_{POTRA1}$ and $Ag43_{N168}$ were performed by mixing the ^{15}N-labelled $TamA_{POTRA1}$ with $Ag43_{N168}$ in a 1:8 and 1:5 molar ratio (highest attainable concentration of $Ag43_{N168}$). Both proteins were dialysed extensively against 50 mM sodium phosphate buffer, pH 6.9 to alleviate the effect of any buffer mismatch on protein chemical shifts. For the demonstration of the paramagnetic relaxation enhancement (PRE) effect between ^{15}N-labelled $Ag43_{N168}$ and MTSSL-labelled $TamA_{POTRA1}$ (Gly31Cys mutant) were

used in 1:3 respectively. HSQC spectra were acquired for ^{15}N-Ag43$_{N168}$ alone, and in the presence of paraTamA$_{POTRA1}$.

In attempts to obtain sequence-specific backbone assignment of Ag43$_{N168}$, 3D ^{1}H-^{15}N HSQC-NOESY (mixing time 400 ms) and 3D ^{1}H-^{15}N HSQC-TOCSY spectra were acquired on 1 mM ^{15}N-labelled Ag43$_{N168}$ in 50 mM sodium phosphate, 100 mM NaCl, 10% (v/v) D_2O pH 6.9 (148). Spectra were processed with AZARA using the MaxEnt method; however, due to poor spectral dispersion and the highly dynamic nature of the protein, only partial assignment could be carried out.

2.17 Protein crystallisation and data collection

For crystallisation, proteins were concentrated (25 mg ml^{-1} for TamA$_{POTRA(22-265)}$ and 15 mg ml^{-1} for the DUF490 construct) using Vivaspin ultracentrifugation spin columns (MWCO 4000-6000 Da) and filtered prior to dispensing into crystallisation trays. Initial crystallisation screens were set up in a 96-well MRC sitting-drop vapour diffusion format (60 μl reservoir solution, 0.5 μl protein + 0.5 μl reservoir) using a Cartesian Honeybee 8+1 dispensing robot. Optimisation of crystal hits was performed using the sitting-drop method with standard optimisation grids, varying precipitant concentration and buffer pH. All crystals were flash-frozen at 100 K prior to X-ray exposure. Where necessary, cryo-protectant was added to the crystals as mentioned in the Results sections. X-ray diffraction data were collected on the i03 and i04 beamlines at the Diamond Light Source synchrotron (Harwell, UK).

2.18 Isothermal titration calorimetry (ITC)

ITC was performed using a VP-ITC instrument (Microcal). All titrations were carried out at 25°C by regular injections of the titrant protein into the chamber containing potential binding partners. The first injection was 1 μl, followed by 28 10 μl injections 180 s apart. Syringe revolution was set to 310 rpm. The heat of dilution of protein into buffer was obtained prior to each sample and subtracted from the reaction. Calorimetric data were

calculated by integrating the area under each peak and thermodynamic values were derived using a non-linear least squares fit and modelled using the best-fit model with Microcal LLC Origin software. In order to minimise buffer mismatch in reactions involving detergent (e.g. Triton X-100) all proteins were buffer-exchanged into an appropriate ITC buffer using a 1 ml HisTrap nickel column and eluted with the ITC buffer + 300 mM imidazole. Imidazole was subsequently removed through extensive dialysis (24-48 hours) against ITC buffer containing detergent.

2.19 Circular dichroism spectroscopy

Protein spectra were measured in 0.1 M sodium phosphate buffer pH 7.5, unless otherwise specified. All spectra were collected in 0.02 cm pathlength quartz cuvettes using a Jasco J-810 spectropolarimeter. Secondary structure estimates were obtained using the CONTIN procedure which is available from the Dichroweb server (149,150).

2.20 Stopped-flow fluorimetry

Pre-steady state $TamB_{DUF490}$-$TamA_{POTRA}$ association kinetics were analysed using an Applied Photophysics Biosequentia stopped-flow spectrofluorimeter. Samples were mixed in a 1:1 (v/v) injection chamber and excited at 280 nm wavelength. Fluorescence was monitored above 305 nm using a cut-off filter, slit width 7 mm. All reactions were performed in 50 mM Tris, 200 mM NaCl, 0.015% (v/v) Triton X-100, pH 7.5 at 22.1°C. The concentration of $TamA_{POTRA}$ was kept constant (5 μM) with a varying concentration of $TamB_{DUF490}$ (20 – 100 μM) under pseudo-first order conditions. The dead time of the instrument was 2 ms, 5000 data points were collected over 10 s and averaged over 10 runs.

2.21 Analytical ultracentrifugation (AUC)

Sedimentation velocity (SV) data were acquired for $TamA_{POTRA}$ and $Ag43_{N168}$ in a Beckman Coulter Optima XL-I analytical ultracentrifuge using an An-50 Ti eight-hole rotor. Protein

and buffer solutions were loaded into 12-mm path-length charcoal-filled epon double-sector centrepieces (360 μl per sample) and spun at 49,000 rpm for ~9 h at 4°C. Scans were collected every 7 min using both interference and absorbance optics (280 nm; a radial range of 5.8 - 7.2 cm, and radial step-size of 0.005 cm were used). SEDNTERP was used to calculate the partial specific volume (from the amino acid sequence of the proteins), the buffer density and viscosity at 20°C and 4°C. Computed values used to calculate the apparent sedimentation coefficients are presented in Table 5.

Table 5 Protein partial specific volumes, buffer densities and viscocities calculated using SEDNTERP for the analysis of SV and SE experiments

Protein	Buffer	Partial specific volume (ml/g)	Buffer density (g/ml)	Buffer viscosity (poise)
$TamA_{POTRA}$	50 mM Tris, 500 mM NaCl, pH 7.5	0.729 (4°C) 0.735 (20°C)	1.022 (4°C) 1.02 (20°C)	0.00157 (4°C) 0.001 (20°C)
$Ag43_{N168}$	50 mM Tris, 100 mM NaCl, pH 7.5	0.705 (4°C) 0.712 (20°C)	1.01 (4°C) 1.0039 (20°C)	0.00156 (4°C) 0.001 (20°C)

Data were analysed using SEDFIT to obtain the apparent sedimentation coefficients using the continuous c(s) distribution model (151). The apparent sedimentation coefficients of $TamA_{POTRA}$ domains were then used to compute the concentration-independent $s^0_{20,w}$ by plotting $TamA_{POTRA}$ concentration versus apparent $s_{20,w}$ and extrapolating the graph to infinite dilution.

Sedimentation equilibrium (SE) data were acquired for $Ag43_{N168}$ at two rotor speeds of 16000 rpm (for monomer) and 10000 rpm (for dimer), with sample concentrations ranging from 0.8-2 mg ml^{-1}. Baseline was determined by accelerating the rotor to 49000 rpm. Both absorbance and interference optics were used as described for SV experiments. Radial step size used was 0.001 cm. 10 scans were taken every 3 hours for 30 hours with a total of 10 scans per speed. Equilibrium data were analysed using SEDPHAT software using species analysis (151).

2.22 Small-angle X-ray scattering

All SAXS data presented in this thesis were collected on the X33 and P12 beamlines at the Deutsches Elektronen Synchrotron (DESY, Hamburg, Germany). Protein concentrations were usually chosen between 0.3-4.4 mg ml^{-1}, unless otherwise specified.

A total of 20 frames were collected for each buffer and sample exposure. Samples were inspected for radiation damage after each exposure. The scattering from buffer alone was acquired before and after each sample and an average of the buffer scattering was subtracted from the sample scattering. Inter-particle interference was inspected by analysing the Guinier region of each scattering curve in the concentration series. Low angle scattering data obtained for the lowest concentration of the protein were merged with the rest of the high angle data from the same protein sample to reduce the contribution of inter-particle effects to R_g determination. All data processing was performed using PRIMUS (152). The distance distribution function, p(r), was obtained by indirect Fourier transform (FT) of the scattering intensity using GNOM (152). A Guinier plot (ln I(q) vs q^2) was used to determine the radius of gyration, R_g. The Guinier-derived R_g was compared with the R_g determined by the indirect FT over the whole scattering range in order to make sure no aggregation was present in the sample. *Ab initio* models of $TamA_{POTRA123}$ were built using DAMMIF and averaged using DAMAVER (153). CRYSOL was used to compute theoretical scattering curves from high-resolution X-ray structures (154). The ensemble optimisation method (EOM) was carried out using the high-resolution model of TamA POTRA domains (PDB ID: 4BZA) taking residues 22-265. Flexible regions were taken to be residues 101-105 and 188-191 and default parameters were used for the program set up (155).

2.23 Hydrodynamic and discrete molecular dynamic (DMD) modelling using US-SOMO

US-SOMO was used to determine hydrodynamic ($s^0_{20,w}$, R_g) parameters based on the high-resolution crystal structures available (156). Discrete molecular dynamics (DMD)

simulations (as part of US-SOMO) were used to explore the conformational space of $Ag43_{N168}$ using the high-resolution model of the Ag43 passenger domain (PDB ID: 4KH3) (130). The Andersen thermostat temperature (T) was set to 0.6 kcal mol^{-1} K^{-1} to allow for sufficient sampling of conformational dynamics around the native state as well as the unfolded state of $Ag43_{N168}$. The run time and PDB output step were adjusted in order to generate 5,000 models. The pool of generated models was refined by a genetic algorithm implemented in the program GAJOE as part of the EOM package (155).

3 Overexpression, purification and structural characterisation of POTRA domains from TamA

Circular dichroism data presented in this chapter has been performed in collaboration with Dr Sharon Kelly

All NMR spectroscopy was performed under the supervision of Dr Brian O Smith

3.1 Introduction

POTRA domains are predominantly found in Omp85 transporter proteins where they mediate intermolecular interactions with partner proteins forming large macromolecular complexes (30,91). Moreover, POTRA domains have been implicated in recognising substrate molecules and facilitating their transport or secretion in a chaperone-like fashion (77,94). The first crystal structures of POTRA domains from BamA were reported by Kim et al and Gatzeva-Topalova et al in different space groups and indicated that these POTRA domains could adopt at least two different conformations, an extended conformation as well as a more compact structure (Figure 3-1) (30,95). Structural comparison of the two different crystallographic models indicated that a potential hinge point exists between POTRA2 and POTRA3, since tandem POTRA1-POTRA2 and POTRA3-POTRA4 domain orientations are conserved between the two models (95).

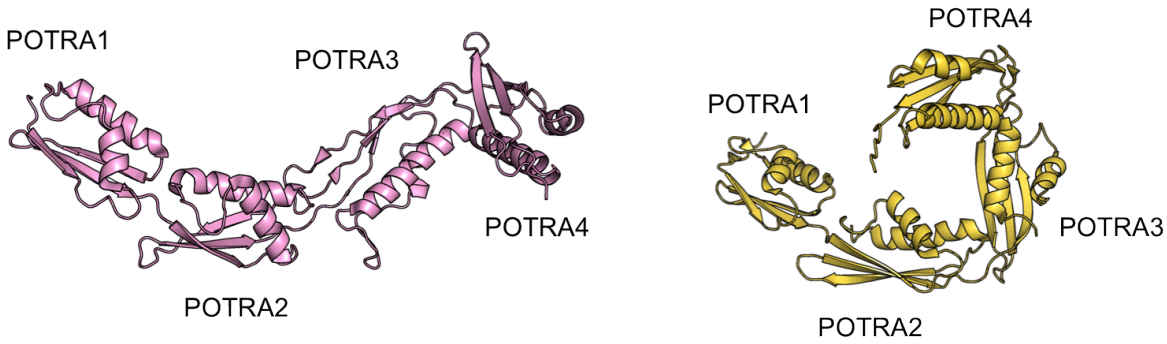


Figure 3-1 Two different conformations of BamA POTRA domains captured *in crystallo.* The orientation of the N- and C-terminal domains is shifted around a hinge point between POTRA2 and POTRA3 from an extended (Left, PDBID: 3EFC) to a more compact (Right, PDBID: 2QCZ) orientation.

The extended conformation of BamA POTRA1-4 domains is in accord with the solution characterisation of the molecule using SAXS, which revealed an extended envelope that could fit all 5 individual POTRA domains (58). However, no analysis of particle flexibility was undertaken; since SAXS reports on the average solution properties of the molecule under investigation, the presence of conformational heterogeneities and ensembles may not be evident if the major population in solution is present as an extended conformer

(58). Another detailed study of the structural conformations of BamA POTRA domains using NMR and SAXS showed that these proteins exist as mixtures of bent and extended conformers in solution, rather than a flexible polypeptide sampling conformational space (95). The biological implications for the existence of this mixture of conformers is currently not understood. Interestingly, disrupting the orientation of tandem POTRA domains in FhaC using a glycine-serine amino acid insertion led to defects in substrate secretion, implying that the relative positioning of the POTRA domains is important for substrate recognition and translocation (106). It is thought that in the context of BamA POTRA domains, substrate association occurs along the exposed edges of the β-sheets and the modulation of domain orientation, either by the substrate or by other members of the BAM complex, would assist in the formation of β-hairpins within translocating substrates.

In parallel with the work presented in this chapter, the structure of residues 22-293 of a TamA POTRA construct possessing residues 22-293 was solved by Dr Justyna Wojdyla and Prof Colin Kleanthous (University of Oxford) who obtained diffracting crystals of TamA POTRA domains, which encompasses its three POTRA domains (22-262) in addition to a C-terminal region (263-293), which in the context of the full-length TamA protein, forms β-strands 1 and 2 of the TamA β-barrel (Figure 3-2A).

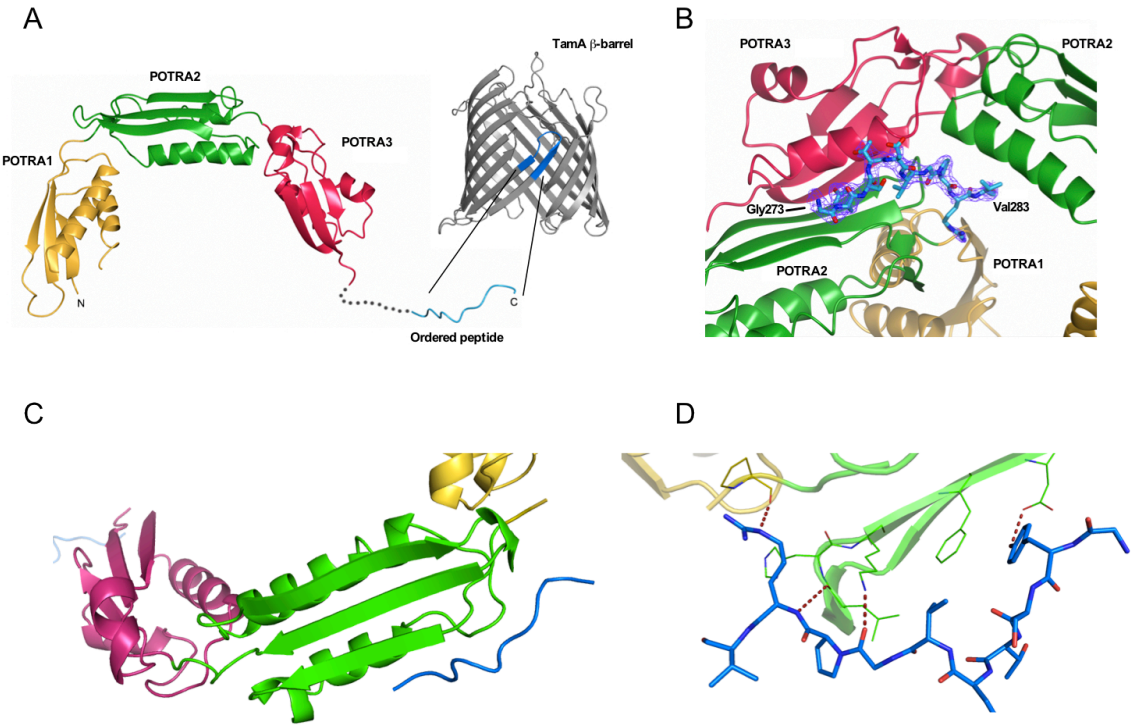


Figure 3-2 (A) Structure of the TamA POTRA domains and origin of the POTRA bound hydrophobic peptide derived from TamA. (B) Electron density of the ordered peptide, which is stabilised by contacts with POTRA2 at the POTRA1-POTRA2 and POTRA2-POTRA3 interfaces of symmetry related molecules. (C) Binding site of the hydrophobic peptide shown in the context of the three TamA POTRA domains with POTRA2 in green. (D) Specific intermolecular contacts between the peptide and TamA POTRA domains emphasises contacts primarily with POTRA2, as well as a part of the POTRA1-POTRA2 linker region.

Interestingly, during refinement of the $TamA_{POTRA(22-293)}$ model performed by Dr Wojdyla, electron density maps showed additional density indicating the presence of a structured peptide derived from C-terminal residues 263-293, which corresponds to the β-hairpin region between the β1 and β2 strands of the TamA β-barrel, of which residues 273-282 could be built into this density. This hydrophobic peptide (GYSTDVGPRV) makes contacts with the POTRA domains of two symmetry-related molecules. One of these interfaces consists primarily of TamA POTRA2 but includes the linker region between POTRA1 and POTRA2 and the other includes both POTRA2 and POTRA3 on the symmetry related molecules (Figure 3-2 B). This finding signifies potential binding regions capable of associating with hydrophobic β-strands derived from β-barrels and a possible evolutionary link to the POTRA domains from BamA that transiently recognise hydrophobic β-strands of OM β-barrel proteins (58). However, in the case of TamA the

interaction of POTRA2 (at the interface with POTRA1) with the C-terminal hydrophobic β-strand may reflect the binding site of the interaction between the POTRA domains of TamA and the DUF490 domain of TamB (Figure 3-2 C,D) (138).

In this chapter the structure and dynamics of the POTRA domains from TamA are characterised using X-ray crystallography and small-angle X-ray scattering. The crystal structure of $TamA_{POTRA}$ domains (residues 22-265, referred to as $TamA_{POTRA(22-265)}$) is reported, in addition to the crystal structure of a different construct (residues 22-293, referred to as $TamA_{POTRA(22-293)}$) solved in collaboration with Prof Colin Kleanthous. During this work the crystal structure of POTRA domains from TamA (residues 22-275) as well as the full-length TamA molecule were published by the Maier group (140). This allowed for a more extensive structural comparison between the 4 available high-resolution models of TamA POTRA domains.

3.2 Results

3.2.1 *Expression and purification of $TamA_{POTRA}$ domains*

Plasmid pPT123 (based on pET21a) was used to express full-length $TamA_{POTRA}$ (residues 22-265 with a C-terminal His_6-tag) without the C-terminal residues present in the crystallised construct (i.e residues 267-293). Cells were grown at 37°C to an OD_{600} of 0.6 and 0.1 mM IPTG was added to induce $TamA_{POTRA}$ production. The temperature was then reduced to 25°C and the cells were then left to grow overnight. Cells were collected by centrifugation, lysed by sonication and the overexpressed protein was purified using Ni^{2+}-affinity chromatography (Figure 3-3A). SDS-PAGE analysis of the eluted fractions showed that overexpressed protein was present mostly in peak 2 (Figure 3-3B). These fractions were pooled and dialysed against 50 mM Tris-HCl, 500 mM NaCl, 1% glycerol, pH 7.5 and purified on a Superdex S75 gel-filtration column. The final purity of the samples was established using SDS-PAGE (Figure 3-3 C,D). The elution volume of $TamA_{POTRA123}$ was approximately 160 ml and the protein eluted as a single peak, suggesting that the molecule behaves as a monodisperse species.

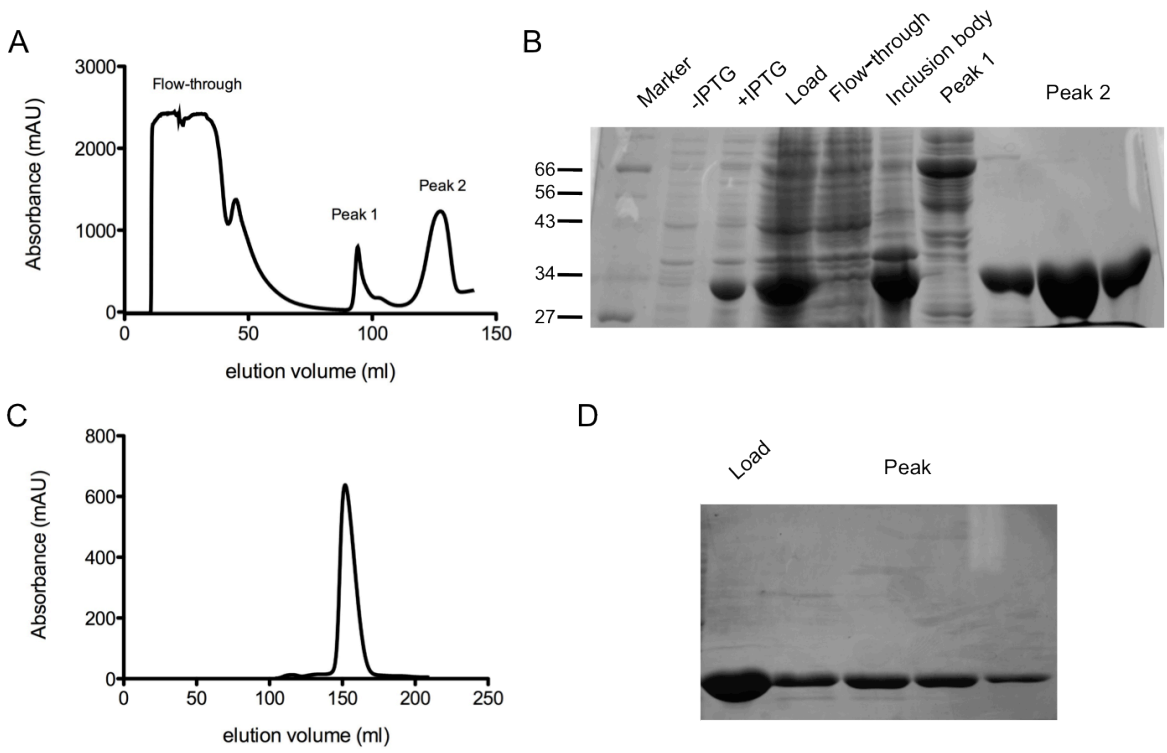


Figure 3-3 Overexpression and purification of $TamA_{POTRA}$ using Ni^{2+}-affinity chromatography and size-exclusion chromatography (SEC). (A) Purification of $TamA_{POTRA}$ using Ni^{2+}-affinity chromatography. (B) SDS-PAGE gel showing post-Ni^{2+} affinity chromatography fractions of His_6-$TamA_{POTRA}$ (30 kDa). Maker M_w are given in kDa. (C) Size-exclusion profile of $TamA_{POTRA}$ run on a HiLoad Supedex S75 column. The protein elutes around 160 ml as a single peak. (D) SDS-PAGE showing purity of $TamA_{POTRA}$ after SEC.

3.2.2 *Solution structure of TamA POTRA domains suggests a more extended particle*

Small-angle X-ray scattering reports on solution properties of the particle, and with the availability of a high-resolution 3D model, comparative analysis can be carried out to determine if the *in crystallo* model is a good description of the solution conformation. SAXS analysis requires almost complete sample monodispersity prior to data acquisition. Therefore, $TamA_{POTRA(22-265)}$ was analysed by sedimentation velocity AUC in order to quantify the number of species present in the purified $TamA_{POTRA(22-265)}$ sample as well as to determine its sedimentation coefficient. A range of concentrations were analysed (9-142 μM), and continuous c(s) distribution analysis revealed that $TamA_{POTRA(22-265)}$ exists as

a single species in solution with no obvious oligomerisation and particle non-ideality (Figure 3-4A). The absolute sedimentation coefficient ($s^0_{20,w}$) of $TamA_{POTRA(22-265)}$ determined was 2.39 S (Figure 3-4B). Computation of the sedimentation coefficient of the high-resolution $TamA_{POTRA(22-265)}$ model using SOMO gave a value of 2.25 S, which agrees well with the experimentally derived value (156). Therefore, $TamA_{POTRA(22-265)}$ exists as a monodisperse particle in solution with overall structural features that correlate well with the high-resolution model derived from X-ray crystallography.

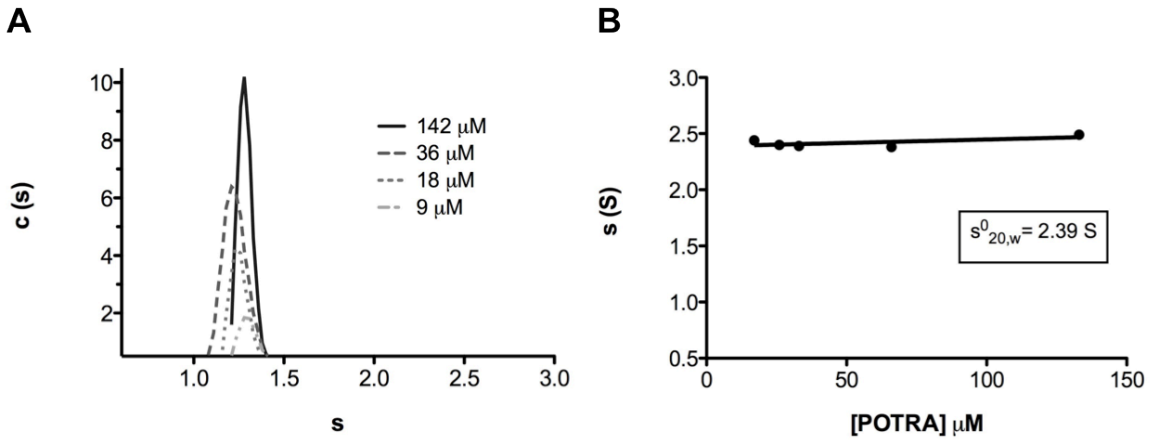


Figure 3-4AUC anaysis of $TamA_{POTRA(22-265)}$ confirms that the particle is monodisperse and ameanable to SAXS analysis. (A) Continuous c(s) distribution at a range of $TamA_{POTRA(22-265)}$ concentrations shows that the molecule is present as a single species in solution with no detectable aggregation in the sample. (B) Extrapolation of calculated $s_{20,w}$ values to obtain the absolute sedimentation coefficient $s^0_{20,w}$.

The high monodispersity of $TamA_{POTRA(22-265)}$ paved the way for further structural characterisation using SAXS which was carried out for a range of $TamA_{POTRA(22-265)}$ concentrations (9-142 μM). No inter-particle interference was observed with varying protein concentration, therefore the data set corresponding to the most concentrated sample was chosen for further analysis (Figure 3-5A). Inspection of the Guinier region showed that the sample was free of aggregate and the measured I(0) gave an estimate of M_w of 26.7 kDa, which is close to the M_w calculated from the amino acid sequence (26.9 kDa). This implies that the sample is monodisperse and the protein is monomeric in solution, which is in agreement with SEC and SV AUC analysis (Figure 3-4). Guinier analysis yielded a radius of gyration (R_g) of 34.4 ± 4 Å (Figure 3-5 B). Calculation of the R_g of the $TamA_{POTRA(22-265)}$ model from the crystal structure using SOMO gave a value of 29.6

Å, which deviates somewhat from the experimentally determined value. This finding indicates that in solution $TamA_{POTRA}$ may exist in a more extended conformation than the crystallographic model suggests.

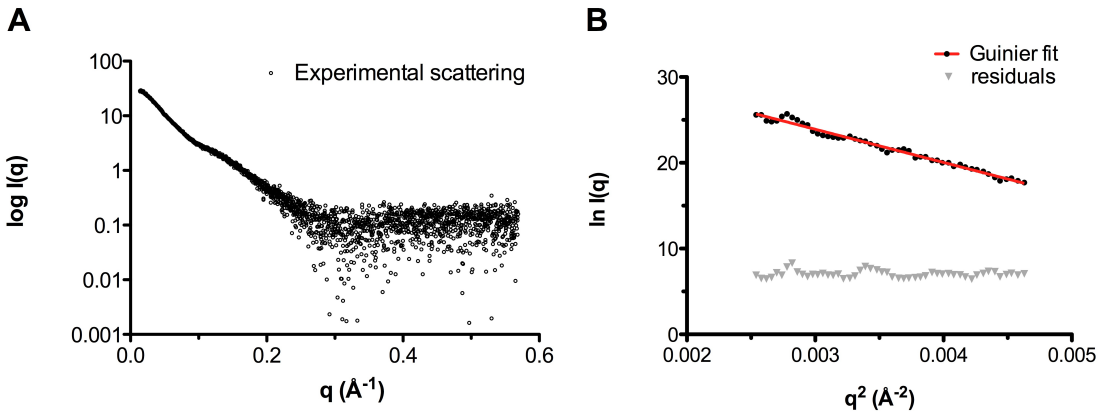


Figure 3-5. SAXS data for $TamA_{POTRA}$ imply an extended particle in solution. (A) Raw scattering curve obtained for $TamA_{POTRA}$ at 4.5 mg ml^{-1}. (B) Guinier fit to the raw scattering data in the Guinier region determines an R_g of 34.4 Å. Linearity of the residuals indicate the absence of inter-particle interference.

Supporting this idea, computation of the scattering curve using the crystal structure of $TamA_{POTRA}$ yields a curve with a poor fit to the experimental scattering data with evident visual deviations, especially at very low angles, reflecting a more compact conformation of the protein *in crystallo* (Figure 3-6).

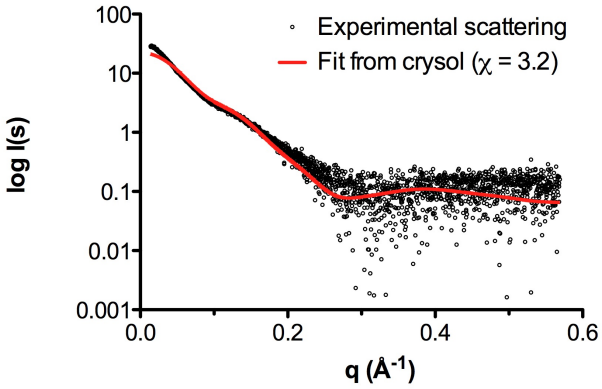


Figure 3-6 Solution structure of $TamA_{POTRA}$ differs from the high-resolution crystal structure. Overlay of experimental scattering and the scattering curve calculated by CRYSOL using the high-resolution crystallographic model shows poor agreement between the particles in solution and the conformation *in crystallo* (PDB ID: 4QAY).

3.2.3 *POTRA domains from TamA exhibit a high degree of flexibility in solution*

Due to the availability of the high resolution structure of $TamA_{POTRA}$, rigid body modelling was undertaken in order to analyse individual domain orientations of POTRA1 and POTRA3 about domain 2 using SASREF (157). This was carried out in order to obtain the representative solution conformations of TamA POTRA domains, consistent with the SAXS data. The second POTRA domain was fixed in space and domains 1 and 3 were rotated and translated about domain 2 coupled with a simulated annealing algorithm minimisation procedure to construct an interconnected model with minimal steric clashes. Several SASREF runs were performed and although the fit of the model with the experimental scattering data improved, the domain orientations were inconsistent between SASREF runs and interconnectivity of the models was poor. This indicates that many solution conformations of $TamA_{POTRA(22-265)}$ domains can describe the experimental scattering data (data not shown). This is contrasted by the BamA POTRA domains which have been shown to exist as rigid body mixtures of bent and extended conformations in solution (95).

Since SAXS provides structural information about the macromolecular flexibility and conformational heterogeneity of a molecule, the molecular flexibility of $TamA_{POTRA}$ domains was inspected (158). The Kratky plot reflects the global overall particle flexibility and is related to the particle volume as well as the scattering contrast between the protein and the buffer components. When the scattering contrast becomes poor, e.g. due to the inherently flexible nature of the molecule, the scattering at high angles increases and results in a tailing effect within the plot. Inspection of the Kratky plot reveals that $TamA_{POTRA}$ possesses a great degree of flexibility, the plot being typical for a protein with multiple distinct domains connected by flexible linkers. Globular, rigid proteins generally have a single peak with low scattering at high angles. However, in the case of $TamA_{POTRA}$ the curve has two peaks with increasing scattering at high angles reflecting the molecule's flexible nature (Figure 3-7A).

Further insights into the flexibility of $TamA_{POTRA}$ domains comes from the inspection of the Porod-Debye plot, which depicts the decay of scattering intensity with the fourth power of q (the momentum transfer) (Figure 3-7B). Typically, for a fully-folded, compact protein a smooth hyperbolic curve with a plateau is observed in a limited region of the scattering curve. However, the loss of this plateau would suggest that the particle is inherently flexible rather than possessing a rigid conformation in solution. In such cases a plateau would be reached under a different (usually smaller) power law that reflects a more rapid scattering decay (159). As evident from Figure 5-5B, no plateau is observed within the Porod-Debye plot of $TamA_{POTRA}$, and the presence of the initial shoulder in the curve implies several scattering contrasts within the sample, which is further emphasised when the data are plotted as $I(q)q^3$ vs q^3 (Figure 3-7C). These scattering contrasts represent markedly different electron density within the average scattering particle and must arise from the conformational heterogeneity of the particle, since the particle was shown to be highly monodisperse using SV experiments and Guinier analysis.

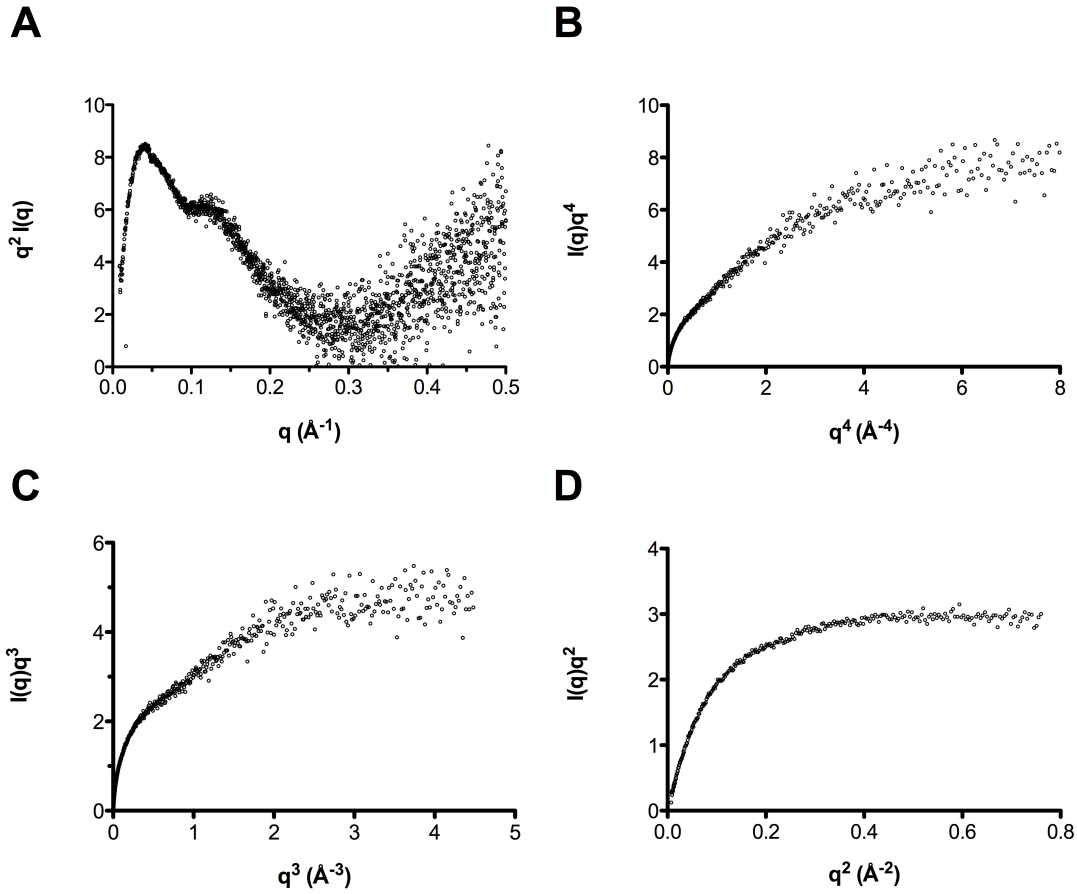


Figure 3-7 Flexibility analysis of the TamA$_{POTRA}$ using scattering intensity decay power laws suggests that the protein is highly flexible. (A) Kratky plot reveals the dynamic character of TamA$_{POTRA}$ in solution since it is indicative of a multidomain protein with flexible linkers. (B) The Porod-Debye plot indicates a complex scattering decay of TamA$_{POTRA}$, which is attributed to its highly dynamic nature. For rigid particles a plateau is usually observed in the range shown (0- 8 x 10^{-4} Å^{-4}). (C) Transforming the scattering intensity as a 3rd power of q reveals a non-hyperbolic shape of scattering decay which emphasises the dynamic nature of the scattered molecule (D) The Kratky-Debye plot shows that the hyperbolic character of the scattering intensity decay is achieved much later (power of 2) than in compact, rigid molecules.

The expected plateau of scattering decay is seen in the Kratky-Debye plot, implying the presence of a flexible, dynamic molecule (Figure 3-7D). Altogether, through the in-depth analysis of scattering intensity decay it can be concluded that TamA$_{POTRA(22-265)}$ exists as a flexible, multi-domain molecule in solution, and this flexibility may be responsible for the extended conformations proposed by the Guinier analysis.

Solution SAXS allows the computation of the maximum dimension (D_{max}) of the particle by performing an indirect Fourier transform of the scattering data. The resultant pair-distance distribution curve directly reflects all the inter-electron distances within the molecule (Figure 3-8A). The distance distribution plot of $TamA_{POTRA(22-265)}$ exhibits two distinct maxima, and a possible third maximum at larger distances reflecting the distinct multi-domain character of the molecule in solution, with an estimated D_{max} of 138 Å. Computation of the D_{max} from the crystal structure yields a value of approximately 90 Å. This stark difference in the maximum molecular dimensions is consistent with the extended conformation of $TamA_{POTRA(22-265)}$ domains in solution in contrast to the crystal structure. *Ab initio* reconstruction of the $TamA_{POTRA(22-265)}$ envelope from the SAXS data indicates that overall the particle envelope is close to that of the crystal structure. However, since this reconstruction reports on the average solution conformation of the molecule, comprising the most consistent dummy atom positions across the 20 computed *ab initio* models, the SAXS envelope may not be directly reflective of the solution conformation, especially since the protein exhibits a high degree of global and local flexibility (Figure 3-8B).

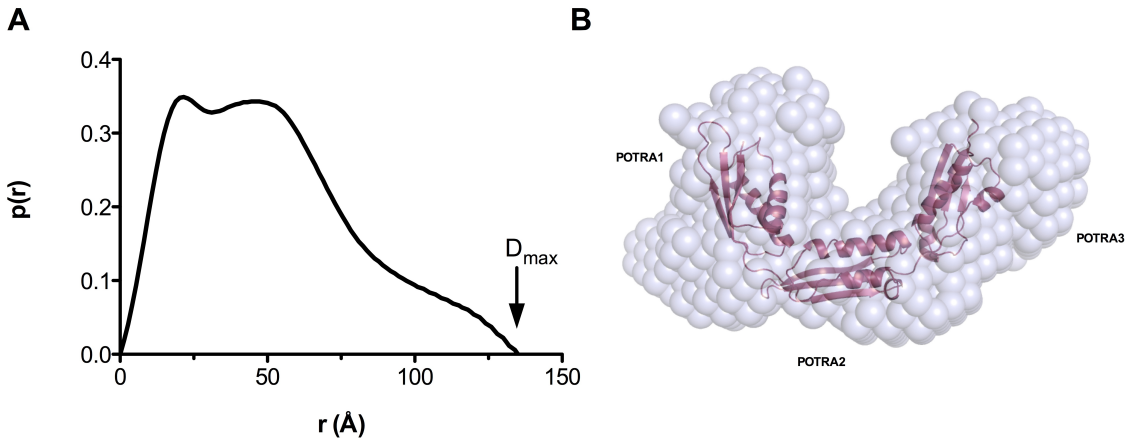


Figure 3-8 Pair-distance distribution and *ab initio* molecular envelope of $TamA_{POTRA}$. (A) The experimentally derived D_{max} of $TamA_{POTRA}$ in solution is 138 Å. (B) Overlay of the *ab initio* molecular envelope of $TamA_{POTRA}$ computed using the experimental scattering with the high-resolution 3D model (PDB ID: 4QAY).

3.2.4 *TamA POTRA domains exist as a conformational ensemble in solution*

The solution structure of $TamA_{POTRA(22-265)}$ was probed further using the ensemble optimisation method (EOM), which attempts to describe the scattering data using ensembles of conformers. The program generates a random pool of 10,000 models of different conformations, calculates a scattering curve for each model and uses a genetic algorithm in order to select ensembles of these molecules that best describe the experimental scattering data (155). High-resolution models of all 3 POTRA domains from TamA were used, and linkers between the domains were assumed to be flexible. The selected $TamA_{POTRA(22-265)}$ ensembles were seen to distribute more towards the extended population with the R_g distribution ranging between 28 and 32 Å, implying that the molecule is somewhat limited with regards to the conformational space that is sampled in contrast to the random pool of conformers (Figure 3-9A). The D_{max} distribution of the selected ensembles suggests that $TamA_{POTRA(22-265)}$ samples both extended as well as more compact conformations in the range of 80-100 Å (Figure 3-9B). Additionally, assuming the presence of $TamA_{POTRA(22-265)}$ ensembles in solution drastically improves the fit of the models to the scattering data compared with the fit obtained with the high-resolution crystal structure (Figure 3-9C). Selected ensembles are presented pictorially in Figure 3-9D. Since only inter-domain dynamics are explored without accounting for the intra-domain loop motions and since the crystal structure lacks the C-terminal His_6 tag the selection of the ensembles is limited by the discrepancy between the high-resolution model and calculated fit of the models with the raw data.

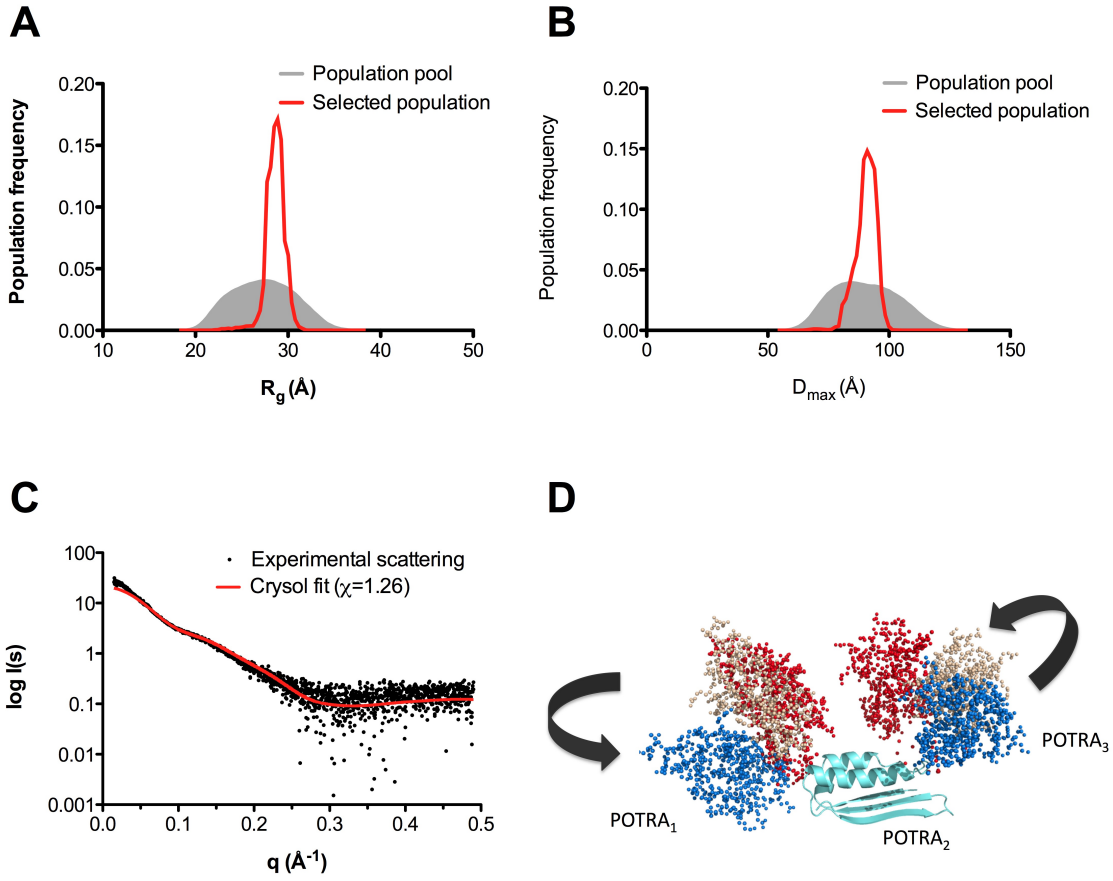


Figure 3-9 EOM analysis of TamA$_{POTRA(22-265)}$ domains suggests that the particle in solution can exist as a subset of ensembles with limited conformational heterogeneity. (A) R_g distribution of the selected ensembles of TamA$_{POTRA(22-265)}$. The selected population tends towards the more extended models from the random pool of conformers (red peak). (B) D_{max} distribution of selected ensemble populations. (C) Overlay of the raw scattering data with the CRYSOL fit of the selected ensembles generated by EOM. The fit is statistically better than that of a single conformation derived from the crystal structure as seen in Figure 3-6. (D) Pictorial representation of the selected conformers of TamA$_{POTRA(22-265)}$ for visual inspection.

3.2.5 *Crystallisation and structural analysis of TamA POTRA domains (residues 22-265)*

In an attempt to capture additional conformations of TamA POTRA domains described by the SAXS data, the construct was subjected to crystallisation screens. Prior to crystallisation screening TamA$_{POTRA(22-265)}$ was concentrated to 25 mg ml^{-1} in 50 mM Tris-HCl, 500 mM NaCl, 5% glycerol, pH 7.5. Initial screening was conducted using the sitting-

drop vapour diffusion method in MRC 96-well plates and consisted of 384 crystallisation conditions. Crystal drops were set up with a 1:1 protein:reservoir buffer ratio, mixing 0.5 µl of protein with 0.5 µl of reservoir and equilibrating the drops against 60 µl reservoir solution at 289 K. Four of these conditions produced a successful hit in the JCSG crystallisation screen, with rod-shaped crystals appearing after 24 hours. The crystals in the 4 different conditions shared the same morphology (Figure 3-10 A).

The crystallisation conditions for the construct $TamA_{POTRA(22-265)}$ differed chemically from the one used for the crystallisation of $TamA_{POTRA(22-293)}$ by Prof. Colin Kleanthous, therefore, one of the conditions was explored to test whether any novel insights into the TamA POTRA structure could be deduced. The crystallisation condition chosen was 1.1 M sodium malonate, 0.1 M HEPES pH 7.0, 0.5% jeffamine M-600. Crystals were flash-cooled in liquid nitrogen and exposed to X-rays on the I24 beamline at Diamond Light Source, Harwell, UK. Crystals diffracted to 2.44 Å and belonged to the space group $P4_12_12$ (Figure 3-10). A complete dataset was collected and processed with iMOSFLM; scaling and merging was done using AIMLESS as part of the CCP4 program suite (160-162). A summary of the processed data is shown in Table 6.

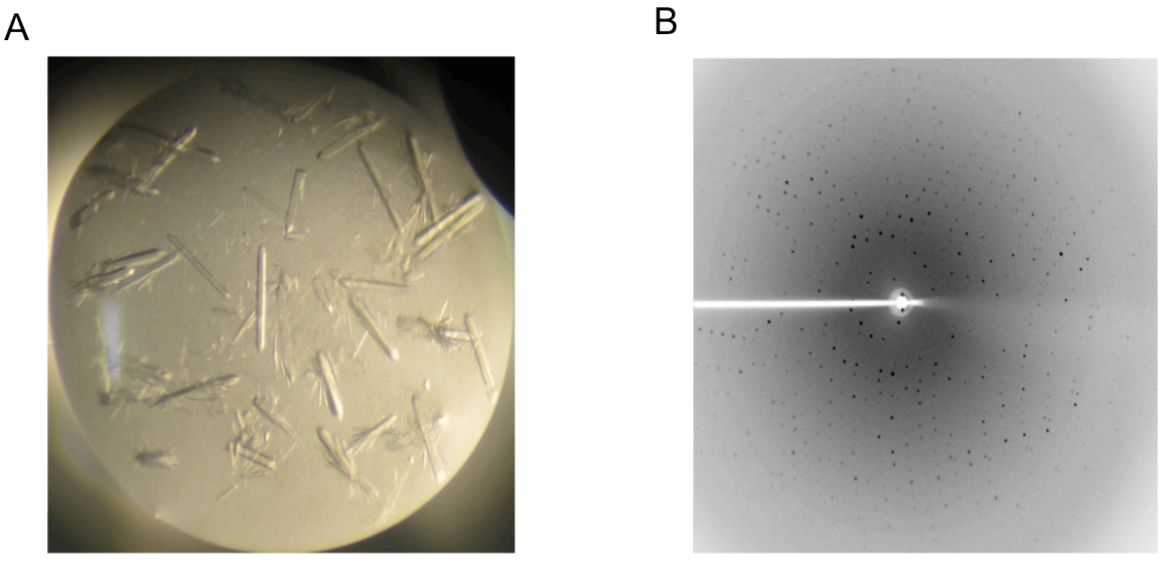


Figure 3-10 Crystallisation and diffraction of $TamA_{POTRA(22-265)}$ crystals. (A Crystals of $TamA_{POTRA(22-265)}$, which belonged to space group $P4_121$. Crystals grew overnight at 289 K. (B) A typical diffraction pattern from $TamA_{POTRA(22-265)}$ crystals. Diffraction extended to 2.44 Å.

Molecular replacement was carried out using the model of $TamA_{POTRA(22-293)}$ from the Kleanthous group (unpublished), using residues 22-265 as a search model in Phaser (163). A single solution was found with LLG scores of 4547.198 and a TFZ score of 13.0, which signified a successful solution. One molecule was found in the asymmetric unit. Several rounds of manual refinement in Coot and REFMAC5 produced a structural model with an R-factor of 0.188 (R_{free} 0.221) suggesting a high confidence in the structure of the model and its fit against the calculated electron density (164,165).

Table 6 Processed diffraction data for the $TamA_{POTRA(22-265)}$ crystal

Diffraction source	DLS I24
Wavelength (Å)	0.9763
Temperature (K)	100
Detector	Pilatus3 6M
Exposure time (s)	0.062
Rotation per image (˚)	0.15
Space group	$P4_12$12
Cell dimensions, *a, b, c* (Å)	78.49, 78.49, 152.80
Resolution (Å)	50.94-2.44 (2.5-2.44)[1]
No. of unique observations	18558 (1345)
Multiplicity	24.9 (25.2)
Completeness (%)	100 (100)
R_{merge} (%)[a]	11.8 (99.7)
R_{pim} (%)[a]	3.3 (28.4)
Mean *I*/sigma (*I*)	29.7 (5.0)

[1] Values in parentheses refer to the highest resolution shell

[a] $R_{pim} = \Sigma_{hkl}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i I_i(hkl)$

[b] $R_{merge} = \Sigma_{hkl}[1/(N-1)]^{1/2}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i I_i(hkl)$ where $I_i(hkl)$ is the *i*th observation of reflection *hkl* and $\langle I(hkl)\rangle$ is the weighted average intensity of all *i* observations of reflection *hkl*

The crystal structure of $TamA_{POTRA(22-265)}$ was structurally identical to the model of the $TamA_{POTRA(22-293)}$ construct with regards to the whole molecule, which suggests that the C-terminal tail does not affect the overall conformation of the protein (Figure 3-11).

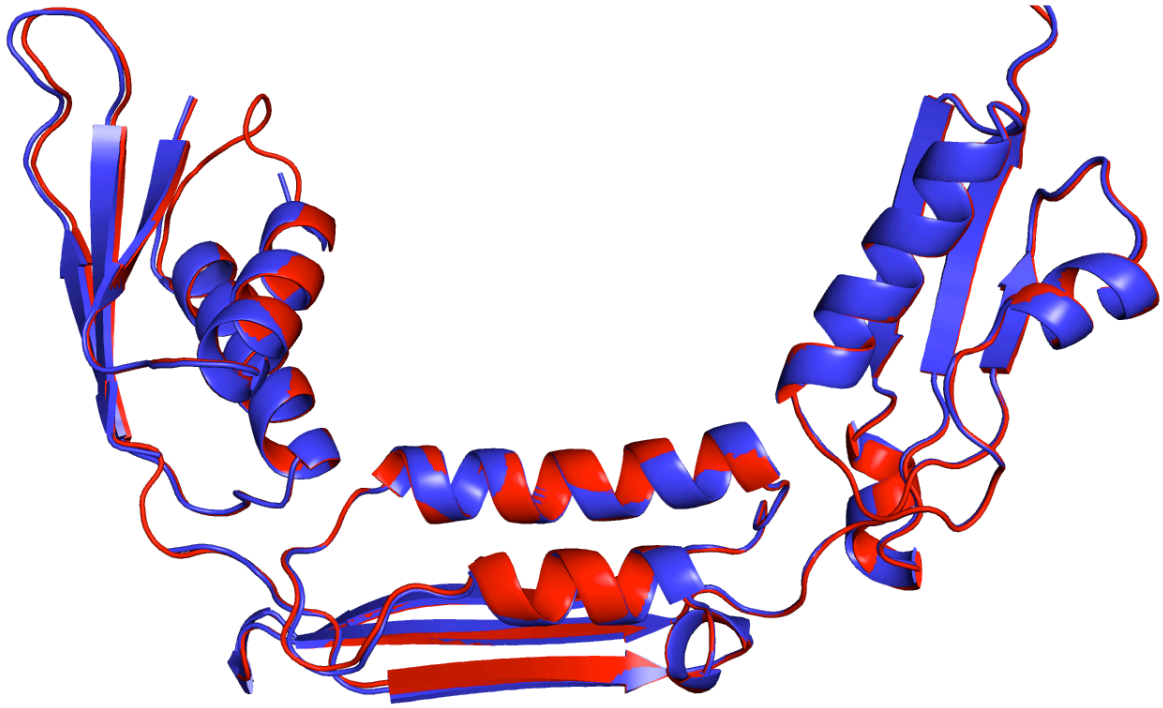


Figure 3-11 Overlay of the two high resolution models of TamA POTRA shows that the structures are essentially identical. Model in red residues 22-265, model in purple 22-293.

The structure of the newly solved $TamA_{POTRA(22-265)}$ model was compared with the available structural models, namely the model of $TamA_{POTRA(22-293)}$ built by Dr. Wojdyla, the model solved by the Maier group (PDBID: 4BZA), as well as POTRA domains from the full-length TamA structure (PDB ID: 4C00), solved by the same group. The overall r.m.s.d. between $TamA_{POTRA(22-265)}$ and $TamA_{POTRA(22-293)}$ was 0.25 Å over the main chain Cα atoms, with an r.m.s.d. of 0.2 Å for the individual domains meaning the overall and individual folds are structurally identical. Alignment of backbone Cα atoms of $TamA_{POTRA(22-265)}$ and TamA POTRA structure from the Maier group yielded an r.m.s.d of 0.45 Å with individual domains overlaying close to those of $TamA_{POTRA(22-293)}$ with the r.m.s.d. of 0.3 Å or less for each individual POTRA domain. Interestingly, the largest Cα deviations were seen for the comparison of the $TamA_{POTRA(22-265)}$ and the POTRA domains from the full-length TamA crystal structure, which could be a result of different crystal packing and molecular lattice contacts. The overall r.m.s.d. between these two models was 1.047 Å and the deviations between the individual domains was 0.54 Å or less. However, on close inspection of the molecular shapes aligned over the middle POTRA2 domain, some degree of domain movement is evident in POTRA domains 1 and 3. This observation suggests that crystal

packing could be influencing the domain orientations, since intra-domain contacts observed in all the crystal structures are not extensive.

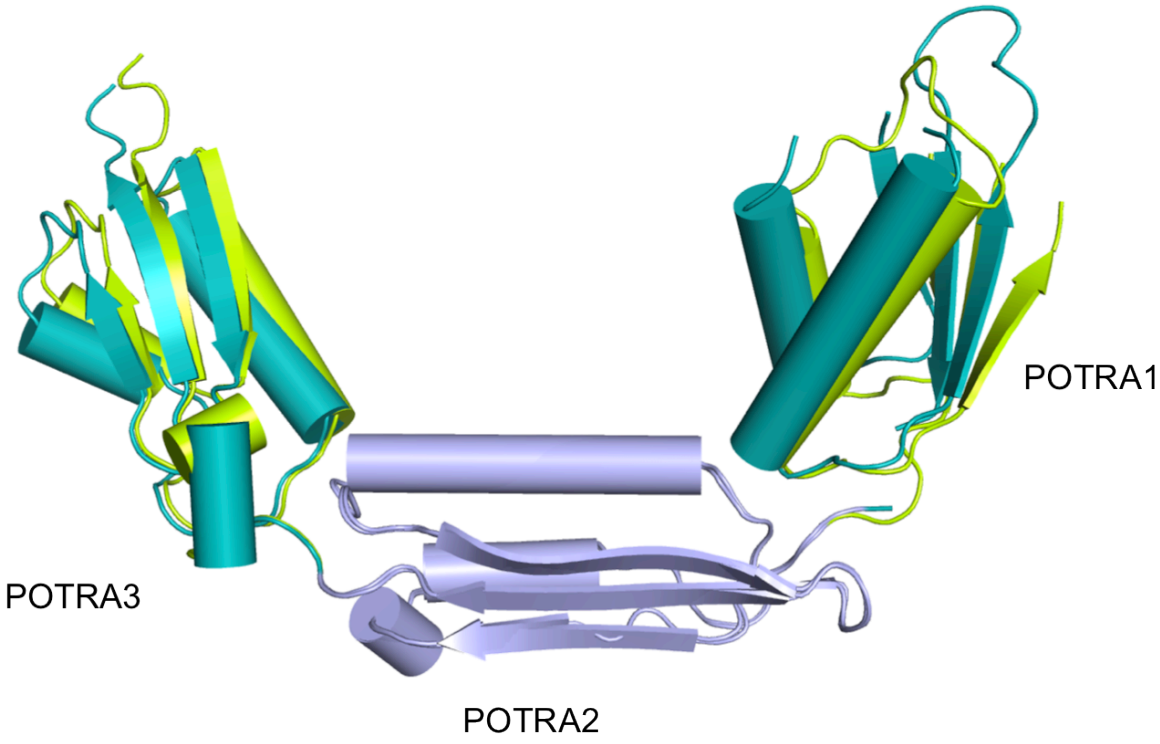


Figure 3-12 Insights into POTRA domain movements. Overlay of the crystal structure of $TamA_{POTRA(22-265)}$ with the POTRA domains from the full-length TamA model (PDBID: 4C00) suggests small domain movements between POTRA1-POTRA2 and POTRA2-POTRA3.

Comparision of B-factors for $TamA_{POTRA(22-265)}$ and TamA POTRA domains from full-length TamA reveal slightly different distributions between the three domains (Figure 3-13). This difference is a result of altered molecular packing in the two crystal types. In the structure of full-length TamA the sole POTRA inter-domain contacts are between the POTRA2 domains of two symmetry-related molecules.

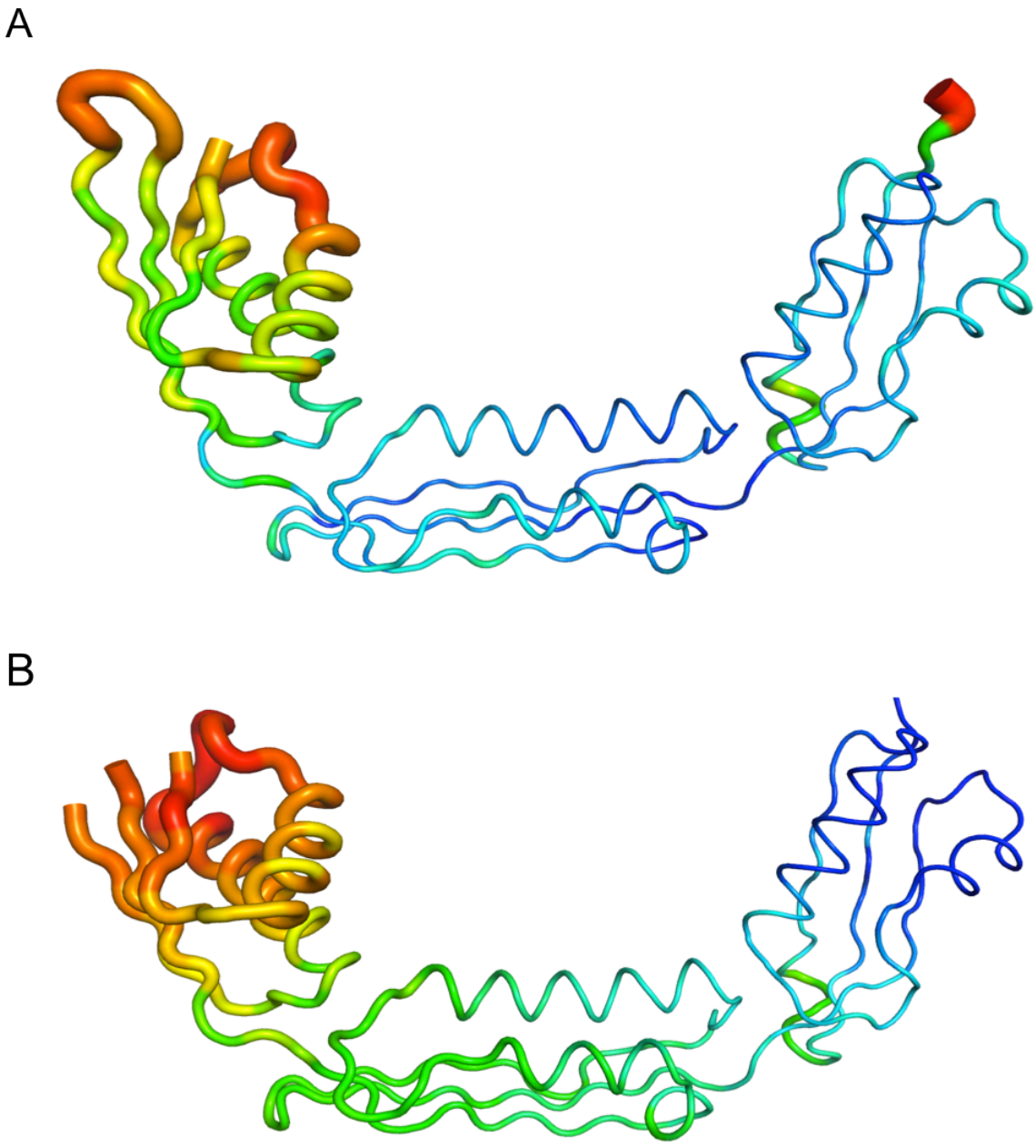


Figure 3-13 Differences in B-factor distributions provide insight into the dynamics of TamA POTRA domains in TamA$_{POTRA(22-265)}$ (A) and POTRA domains from full-length TamA model (B). Dark blue suggests rigid backbone with a shift to more dynamic backbone represented by green to yellow to red.

In contrast, the crystal structures of TamA$_{POTRA(22-265)}$ and TamA$_{POTRA(22-293)}$ reveal extensive contacts between POTRA domains 2 and 3 as well as a weak contact between domains 1 and 2 which explains the lower B-factor distribution in POTRA2 and POTRA3. In order to alleviate the contribution and constraint of crystal packing on the B-factor profiles of TamA POTRA models, a more thorough analysis was carried out. Only main-chain B-factors were considered and a normalisation to the main chain B-factor was carried out

for each model using the Phenix software package (166). In all instances it was evident that POTRA1 possesses higher main-chain B-factors than the other two domains, which suggests that this domain exhibits the highest flexibility within the structure. No differences in B-factor profiles were seen between $TamA_{POTRA(22-265)}$ and $TamA_{POTRA(22-293)}$ (Figure 3-14A). Inspection of the B-factor profile for the models of $TamA_{POTRA(22-265)}$ and TamA POTRA domains from the Maier group indicate that the largest differences in intra-domain dynamics lie within the first POTRA domain (Figure 3-14B). However, no differences were seen in the other two POTRA domains. Interestingly, and again due to differences in the crystal packing, noticeable dynamic differences can be seen between $TamA_{POTRA(22-265)}$ and POTRA domains derived from the full-length TamA crystal structure in terms of their B-factor distribution. The third POTRA domain in $TamA_{POTRA(22-265)}$ has a more dynamic character than POTRA3 domain of the full-length TamA model. This could be the result of extensive hydrophobic contacts formed between the third POTRA domain and the periplasmic loops of the TamA β-barrel, as opposed to the β-augmentation observed between POTRA2 and POTRA3 domains in $TamA_{POTRA(22-265)}$ model. Additionally, POTRA2 in the TamA structure has elevated B-factors compared with POTRA2 of $TamA_{POTRA(22-265)}$ model. Altogether, this analysis reveals that POTRA domains from TamA exhibit variable dynamic profiles that are influenced by the packing of the molecules within the crystal lattice.

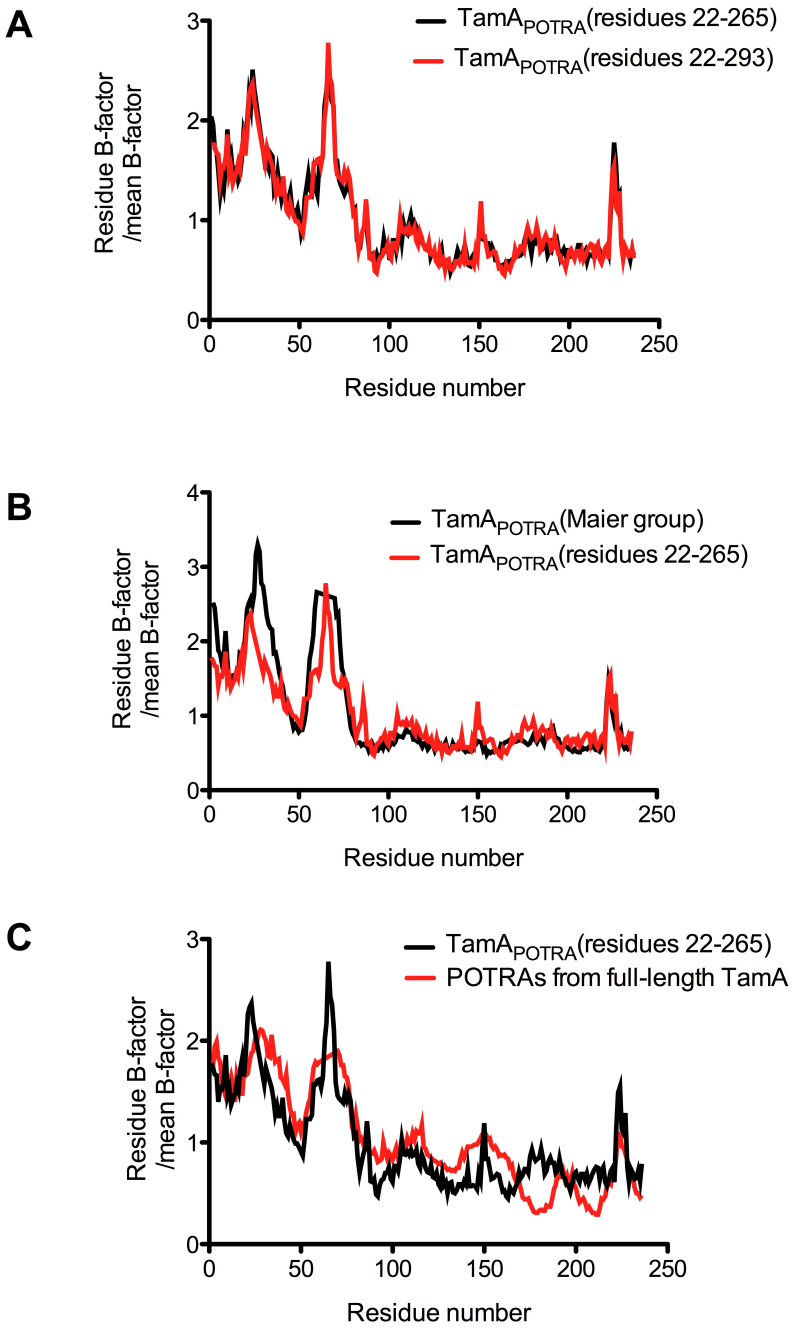


Figure 3-14 B-factor profiling of models of POTRA domains from TamA reveal the putative dynamic character of these molecules. Prior to analysis, all the PDB chains were renumbered to start with residue 1, however in reality they begin at residue 22, after the signal peptide sequence. This was done for comparative reasons solely**.** Only main-chain B-factors are taken into account and are normalised to the mean B-factor calculated over the whole structure used for comparison. (A) Plot of normalised B –factors between $TamA_{POTRA(22-265)}$ and $TamA_{POTRA(22-293)}$ (B) Normalised B-factor plot of $TamA_{POTRA(22-265)}$ and the structure of TamA POTRA domains solved by the Maier group. (C) Normalised B-factors plotted for $TamA_{POTRA(22-265)}$ against the POTRA domains from the structure of the full-length TamA molecule.

3.2.6 *NMR and CD spectroscopic analyses indicate a molten-globule-like character of POTRA2 in TamA$_{POTRA12}$ construct*

Further structural characterisation of TamA$_{POTRA}$ domains was attempted using NMR spectroscopy. Initially, individual POTRA domains from TamA were overexpressed. TamA$_{POTRA1}$ could be readily purified in large quantities. However, attempts to purify TamA$_{POTRA2}$ and TamA$_{POTRA3}$ in isolation were unsuccessful as these proteins accumulated in inclusion bodies and despite several attempts could not be successfully purified. TamA$_{POTRA}$ is inherently prone to aggregation in simple buffers lacking high concentrations of NaCl and glycerol, which makes this construct unsuitable for use over extended time periods at high concentrations and temperatures required for extensive NMR analyses. Therefore, TamA$_{POTRA12}$ and TamA$_{POTRA1}$ were used for the structural studies and ^{15}N-labelled TamA$_{POTRA1}$ and TamA$_{POTRA12}$ were produced. The ^{1}H-^{15}N HSQC spectrum of TamA$_{POTRA1}$ showed a good chemical shift dispersion over a large ^{1}H range and a uniform cross-peak linewidths indicating that TamA$_{POTRA1}$ was folded and stable. However, inspection of the HSQC spectrum of TamA$_{POTRA12}$ indicated that although POTRA1 was also well-folded and stable, with chemical shift positions almost identical to those of TamA$_{POTRA1}$, the chemical shifts originating from POTRA2 were poorly dispersed with non-uniform cross-peak linewidths. Scanning a range of temperatures from 293-308 K as well as additives such as L-Arg and L-Glu commonly used to improve protein stability during NMR studies did not lead to spectral improvement. These analyses suggested that the second domain likely exists in a partially unstructured or molten globule state in solution (Figure 3-15).

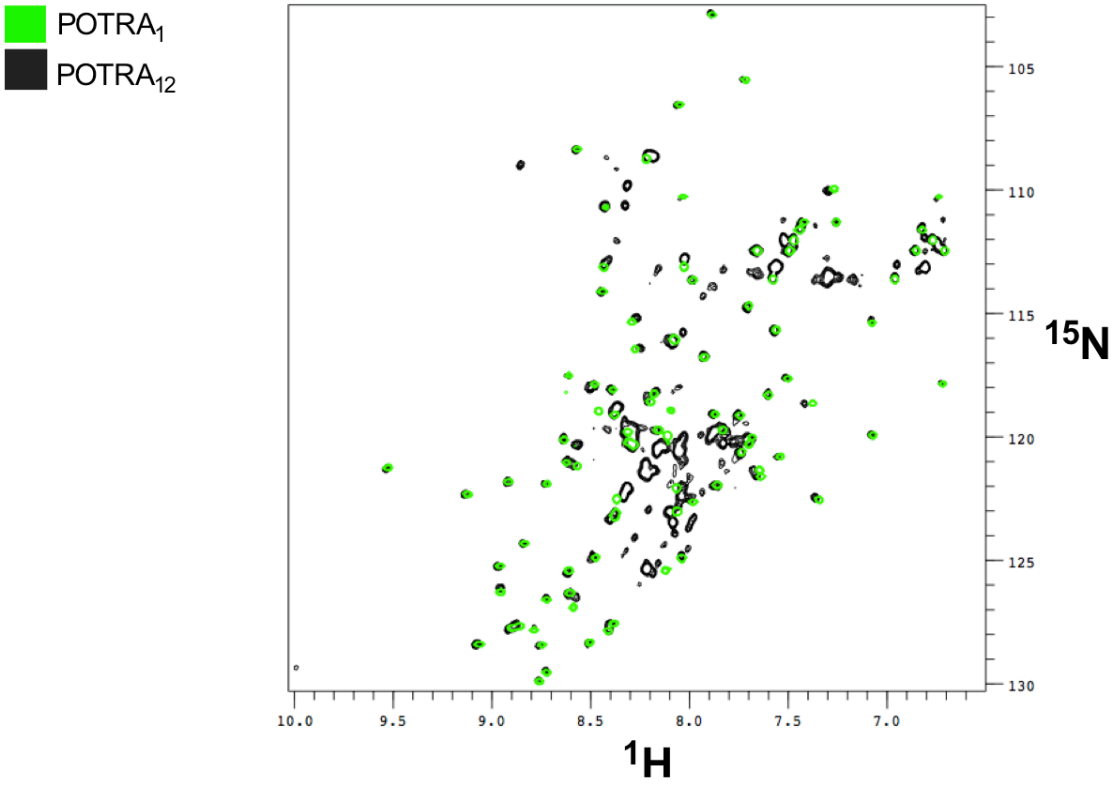


Figure 3-15 NMR spectroscopy shows that POTRA2 exists in a molten globule-like conformation. Overlay of HSQC spectra of ^{15}N-labelled TamA$_{POTRA1}$ and TamA$_{POTRA12}$ reveals that POTRA2 exists as a molten globule in solution due to poor chemical shift dispersion arising from the cross-peaks from domain 2. Good chemical shift dispersion of POTRA1 suggests that it is folded. Small shifts in cross-peaks for POTRA1 versus POTRA12 are attributed to differences in the experimental conditions in which the spectra were acquired.

In order to analyse the secondary structure of TamA$_{POTRA12}$, CD spectroscopic analysis was undertaken. Far-UV CD measurements indicate that at least 50% of the protein possesses regular secondary structure, with 31.6% α-helix, 28.4% β-strand and 40% random coil (Figure 3-16A). These measurements are in accord with the calculated secondary structure elements based on the high resolution model of TamA$_{POTRA}$ suggesting that despite poor chemical shift dispersion for POTRA2, the domain still possesses all of its regular secondary structure elements.

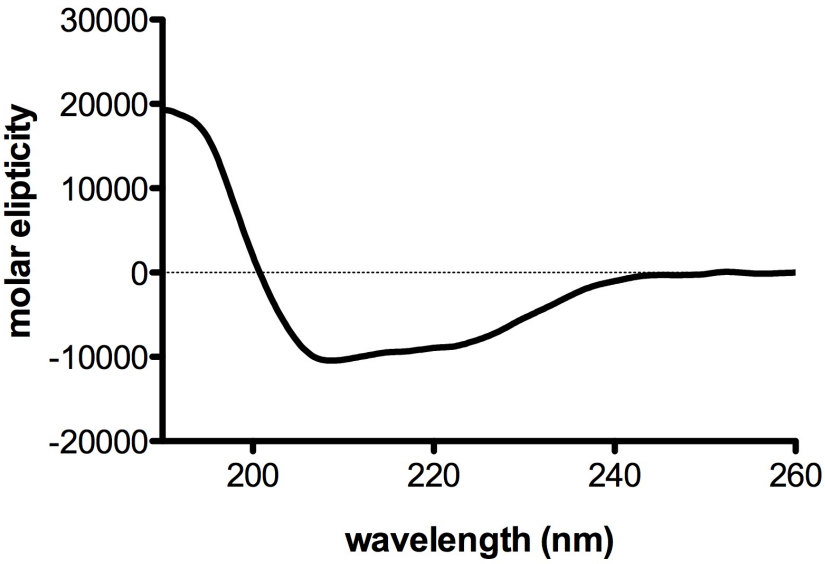


Figure 3-16 Circular dichroism spectroscopy indicates that $TamA_{POTRA12}$ possesses all of its secondary structure despite poor NMR spectra. Far-UV spectrum of $TamA_{POTRA12}$ (36 μM) in 0.05 M sodium phosphate, pH 7.5 shows that at least 50% of the polypeptide is composed of secondary structure elements. Estimates are based on the average of all matching solutions and calculated using the Provencher and Glockner method (150).

Based on the NMR and CD spectroscopic data it can be concluded that POTRA2 could exist as a molten-globule-like in solution retaining all of its regular secondary structure, yet lacking a well-defined tertiary fold.

3.3 Discussion

Extensive structural studies of POTRA domains from BamA suggest that these molecules exhibit complex structural and biophysical dynamics (30,58,95,96,167). The biological implications of this structural diversity are currently not understood. In this chapter, TamA POTRA domains were investigated using several structural solution techniques and X-ray crystallography. The high-resolution crystal structure of the three POTRA domains from TamA is structurally identical to the crystal structures solved by Dr Wojdyla and Prof. Kleanthous (University of Oxford) as well as that reported by Gruss *et al (*140). All these models overlay with the POTRA domains from the crystal structure of full-length TamA, yet minor displacement of POTRA1 and POTRA3 domains between the models is

evident. As with POTRA domains from BamA, where only the extended conformation of these domains is captured, the structural analysis of TamA POTRAs using SAXS suggests that in solution these domains present a more complicated conformational behaviour than suggested by the crystallographic studies. In contrast to the BamA POTRA domains, which have been shown to exist as mixtures of bent and extended conformations, TamA POTRAs exhibit a higher degree of flexibility and behave as ensembles of different, but restricted, conformations. This structural fluidity could have important implications for the function of the TamA molecule in the context of the full protein. Crystal structures of BamA have shown that the POTRA domain closest to the β-barrel is able to change its orientation relative to the barrel pore, and thus could serve as a plug, the position of which could be modulated by an incoming substrate or another member of the BAM complex that associates with it, e.g. BamD (30).

Further insights into POTRA domain conformations in the context of the full TamA molecule are needed. These could be obtained using, for example, small-angle neutron scattering in combination with molecular dynamics. A similar approach was recently been undertaken to investigate the solution structure of FhaC, which revealed that in solution, like in the crystal structure, the barrel pore is likely to be closed by the N-terminal α-helix H1 (168). Therefore, global conformational changes must occur in FhaC for this helix to be ejected from the barrel during substrate translocation. Alternatively, the structures of TamA POTRA could be probed using PELDOR spectroscopy and the site-specific paramagnetic spin labelling approach to look at distance distributions of paramagnetic probes in these domains in order to generate a clearer picture of domain-domain orientations (167). Other approaches such as residual dipolar coupling and paramagnetic relaxation enhancement NMR methods, which probe the structure and dynamics of molecules, would not be amenable to the study of TamA POTRA conformations due to protein instability and poor spectral dispersion, coupled with extensive experiment times.

3.4 Conclusions

In this chapter the crystal structure of $TamA_{POTRA22-265}$ presented gives some insights to the dynamic nature of TamA POTRA domains through structural comparisons with other TamA POTRA high-resolution models. The conformation of TamA POTRA domains was also described through SAXS and revealed that in solution these domains exist as more extended, flexible molecules which sample a limited conformational space. Unexpectedly, through the use of NMR and CD spectroscopy it was concluded that $TamA_{POTRA12}$ exists in a partial molten-globule like state, with POTRA1 being fully folded and POTRA2 lacking any tertiary structure.

4 Structural and biophysical characterisation of DUF490 domain from TamB

Circular dichroism data presented in this chapter has been performed in collaboration with Dr Sharon Kelly

4.1 Introduction

TamB is a large inner membrane protein that has been shown to form a periplasm-spanning complex with an OM protein TamA (138). The existence of TamB in the same operon as TamA suggests an evolutionary and functional conservation of the protein pair in the biogenesis and secretion of bacterial proteins. The *tamAB* operon is highly conserved amongst Gram-negative bacteria signifying that both proteins are involved in some specialised biological process, recently demonstrated to be AT passenger domain secretion across the OM (138). This is contrasted by the hetero-oligomeric BAM complex, where the members BamABCDE are not encoded in an operon together but function as a complex that is responsible for the biogenesis of bacterial OMPs or overall complex stability. Although *in vitro* studies have shown that BamA alone is sufficient for the incorporation of unfolded OMPs into phospholipid vesicles, other components of the BAM complex act as accessory factors that assist BamA functionally, with BamD being essential for bacterial cell survival (169). It is therefore inviting to speculate that TamB acts as an accessory factor to TamA, contributing to the process of AT passenger domain transport and secretion across the OM. Previous bioinformatic analyses classified TamB as a putative β-barrel protein associated with OM integrity and LPS assembly (170). However, through extensive biochemical characterisation it has been demonstrated that TamB is in fact an inner membrane-bound protein. Protease-shaving experiments indicate that at least part of TamB is periplasmic, implying that some of the protein could be located in the cytoplasm or is protease-resistant (138). Moreover MorC, a distant homologue of TamB present in *Aggregatibacter actinomycetemcomitans*, a periodontal pathogen, is involved in OM morphology biogenesis and secretion of leukotoxin, which is a haemolysin-like toxin. Interestingly, MorC is found in an operon with two other genes, *omp67* which is a distant TamA homologue, and *ppx* which is an exopolyphosphatase that hydrolyzes polyphosphate molecules involved in a wide range of biological processes in bacteria (171). The greatest sequence conservation between TamB and MorC lies within the conserved C-terminal domain of unknown function, DUF490, which is part of the Pfam04357 family of proteins (172). The C-terminal DUF490 domain from TamB has been implicated in the formation of the cell envelope-spanning complex with TamA. However,

to date no structural or biochemical information is available for this domain. In this chapter the overexpression, purification and preliminary structural characterisation of DUF490 is presented.

4.2 Results

4.2.1 *Expression, refolding and purification of the $TamB_{DUF490}$ domain*

$TamB_{DUF490}$ was expressed from a pET28a plasmid with an N-terminal His_6-tag in *E. coli* BL21(DE3) cells and purified from inclusion bodies (IBs). Cells were grown at 37°C to an OD_{600} of 0.4-0.6 and induced with 1 mM IPTG for 4-5 hours, pelleted and lysed by sonication. $TamB_{DUF490}$ was then purified from IBs. IBs were washed with 50 mM Tris-HCl, 50 mM NaCl, 0.1% Triton X-100, pH 7.5, followed by several washes with the same buffer without Triton X-100. IBs were then solubilised in 50 mM Tris-HCl, 50 mM NaCl, 5% sodium N-lauroyl sarcosinate and left at room temperature for 2 hours. Solubilised IBs were then diluted (1:10) in refolding buffer (50 mM Tris-HCl, 200 mM NaCl, 0.3 M L-arginine, 0.1% Triton X-100, pH 8.0) and dialysed for 24 hours at 4°C in refolding buffer, followed by an additional 24 hour dialysis in 50 mM Tris-HCl, 200 mM NaCl, 0.015% Triton X-100. Refolded protein was further purified using Ni^{2+}-affinity chromatography and SEC and assessed by SDS-PAGE for purity (Figure 4-1 A,B).

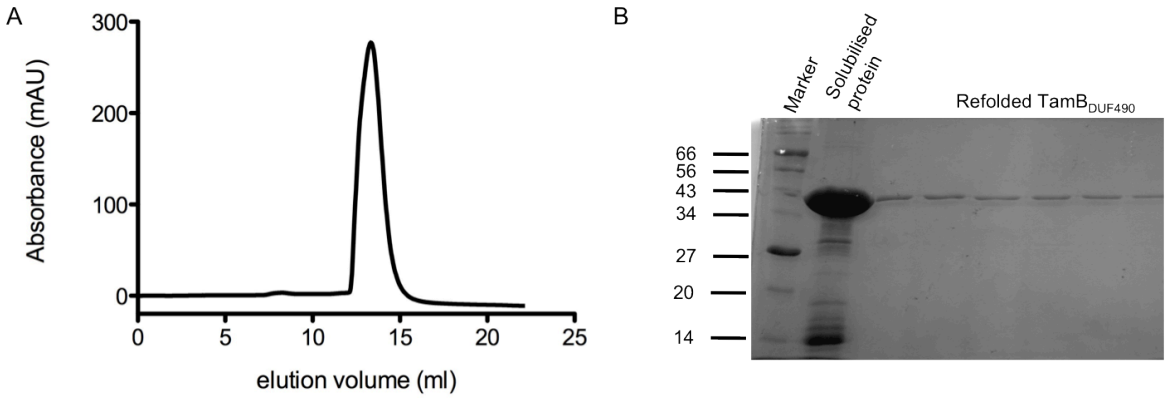


Figure 4-1 Overexpression and purification of $TamB_{DUF490}$. (A) SEC profile of $TamB_{DUF490}$. The elution volume of $TamB_{DUF490}$ is approximately 13 ml on an analytical Superdex S200 column, which corresponds to the refolded protein (42 kDa) including the Triton X-100 micelle (~95 kDa). (B) SDS-PAGE of refolded $TamB_{DUF490}$ (42 kDa), solubilised from inclusion bodies. M_w markers are in kDa.

The monodispersity and molecular weight of $TamB_{DUF490}$ was analysed by analytical SEC using a Superdex S200 GL10/300 column. The analytical SEC column was calibrated with known standards ranging from 29 – 669 kDa. Comparison of the elution profile of refolded $TamB_{DUF490}$ against the SEC standards suggests that the particle elutes at a position corresponding to a molecular mass of 150 kDa (Figure 4-2). The calculated molecular weight of monomeric $TamB_{DUF490}$ based on its amino acid sequence is 42 kDa, and the size of a Triton X-100 micelle is estimated to be around 95 kDa (173), suggesting that $TamB_{DUF490}$ may be present as a complex consisting of monomeric $TamB_{DUF490}$ embedded in a Triton X-100 micelle. However, it cannot be ruled out that $TamB_{DUF490}$ is oligomeric and does not strongly associate with Triton X-100 micelles. Static light-scattering as well as SEC multi-angle laser light scattering experiments have shown that membrane and detergent-bound proteins behave as complexes of proteins and micelles with additive molecular weights, the relative elution volumes of these proteins alone may be less informative about their overall oligomeric make up (174,175).

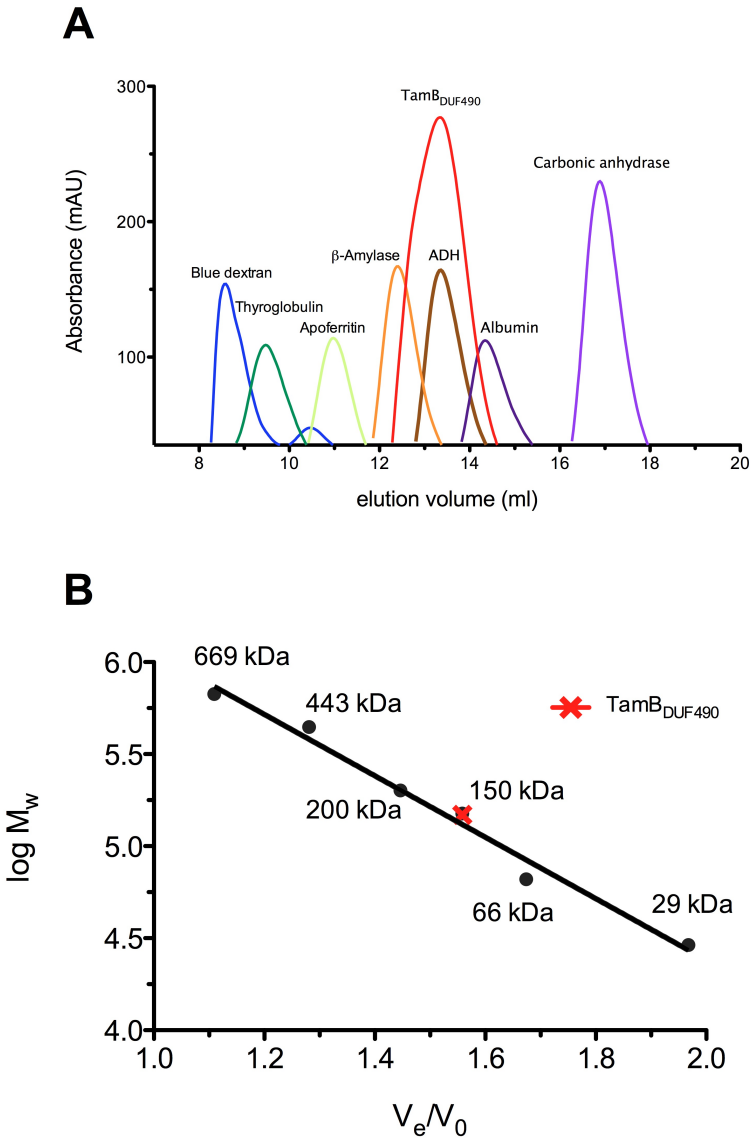


Figure 4-2 Analysis of refolding of TamB$_{DUF490}$ by analytical SEC. (A) The elution profile of TamB$_{DUF490}$ against several standards used for Mw estimation. TamB$_{DUF490}$ elutes around 13 ml. (B) Estimation of TamB$_{DUF490}$ Mw based on elution volumes of the standards using a standard curve. Mw was calculated to be 150 kDa, which is thought to correspond to TamB$_{DUF490}$ (42 kDa + Triton X-100 micelle, ca 95 kDa). A Superdex S200 GL10/300 analytical SEC column was used during this analysis. The buffer was 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 0.015% Triton X-100.

4.2.2 *Sequence-based analysis of DUF490 domain from TamB*

Since very little structural and biochemical information is available about the DUF490 domain from TamB, sequence-based methods were utilised in order to investigate

predicted structural features of this domain. Sequence-based secondary structure prediction of DUF490 from TamB using Jpred suggests that this domain is predominantly β-sheet rich, with three putative α-helices located an the C-terminal end of the domain (Figure 4-3) (176). Analysis of the amino acid sequence of the DUF490 domain using ThreaDom suggests that DUF490 functions as a single domain rather than a multi-domain polypeptide (177). ThreaDom uses multiple threading-based algorithms in order to predict putative domain boundaries and domain discontinuities, with the assumption that residues within the domain core are generally more conserved than the inter-domain linkers and that homologous domains possess similar domain structures (177). Additionally, inspection of the amino acid sequence of DUF490 using several protein disorder prediction programs such as DisEMBL™ and PrDOS does not reveal any systematic presence of long, unstructured regions within the protein (data not shown) (178,179).

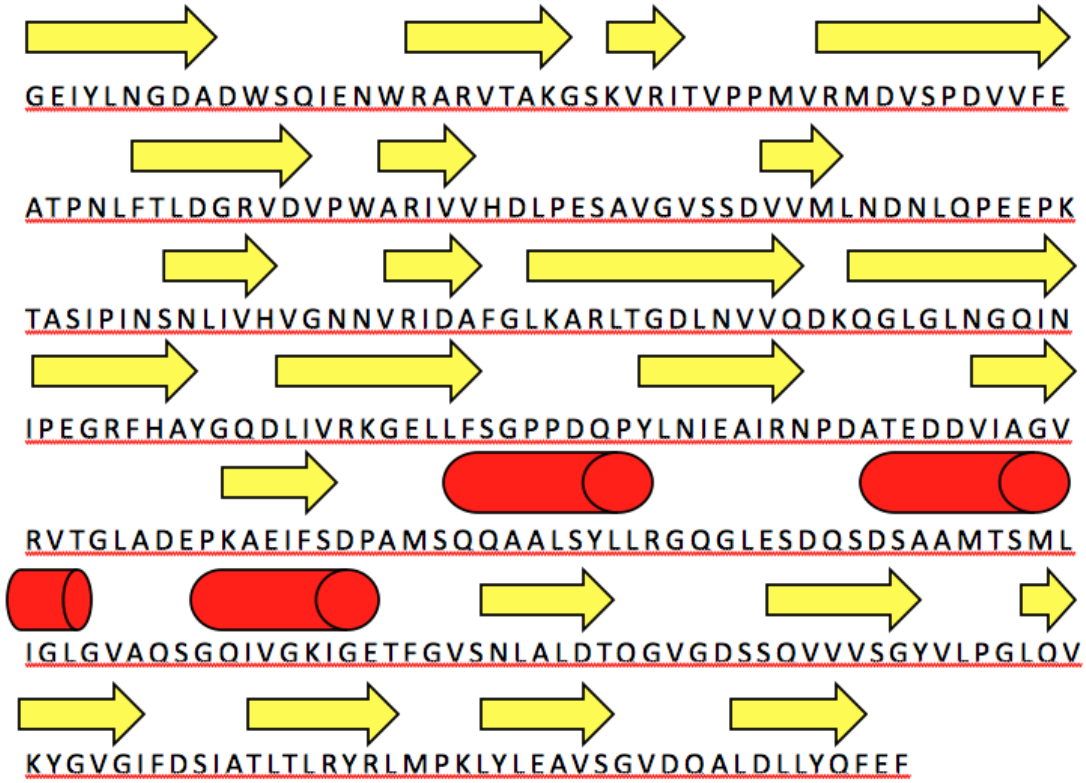


Figure 4-3 Secondary structure prediction of DUF490 using Jpred indicates a mixed α/β fold. Reliability for the secondary structure prediction was high, most secondary structure stretches having scores of greater than 7 (on a 0 to 9 grade). Arrows represent putative β-strands, cylinders represent putative α-helices.

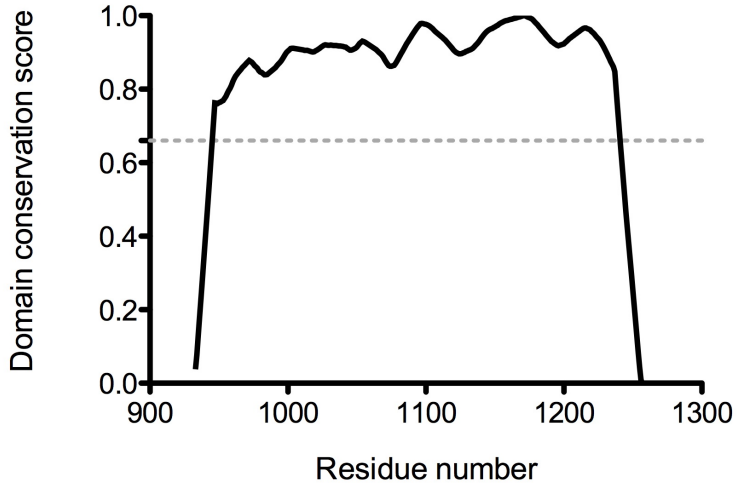


Figure 4-4 Analysis of DUF490 sequence by ThreaDom suggests no discontinuities within the conserved domain. The calculated discontinuity cut-off was 0.66 (grey dashed line).

Interestingly, TMpred analysis of the whole DUF490 domain produces a high positive score for the presence of a putative transmembrane helix in the predicted α-helical segment with an N-terminal inside-to-outside topology (Figure 4-5A) (180). This would imply that the C-terminal 90 amino acids of TamB are located in the periplasmic space whilst the rest of the DUF490 domain is cytoplasmic (Figure 4-5B).

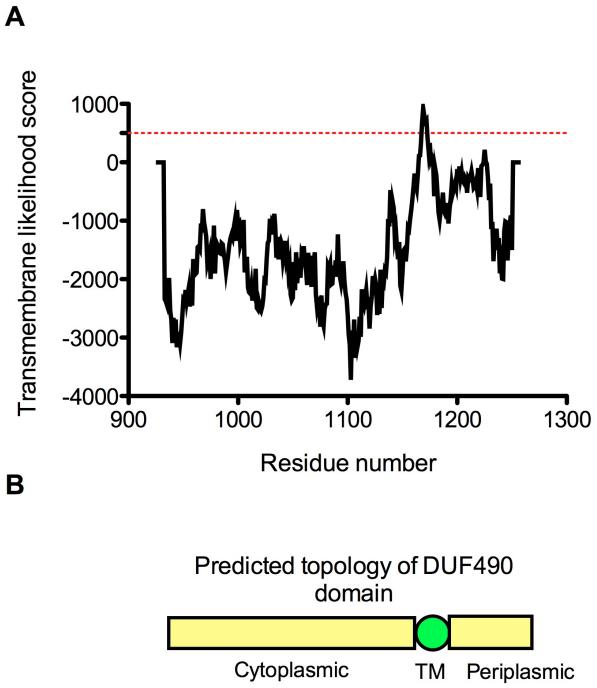


Figure 4-5 TMpred analysis of DUF490 domain of TamB reveals the presence of a putative transmembrane helix within the C-terminal domain. (A) Only scores above 500 (dashed red line) are considered significant. Only the strongly preferred model of inside-to-outside topology is shown. The predicted transmembrane region lies between residues 1161 and 1179, with a maximum score of 927. (B) Diagrammatic representation of DUF490 domain according to the TMpred prediction.

Structural homology modelling was attempted for the DUF490 domain of TamB using Phyre and iTasser; however, due to very low sequence homology to any of the structural models available in the PDB database, the confidence in the models was very low (data not shown) (181,182).

4.2.3 *Expression, purification and characterisation of the C-terminal fragment of $TamB_{DUF490}$ (residues 1180-1259)*

Previous studies have shown that the last 80 amino acids (residues 1180-1259) of TamB are necessary for TAM complex formation (138). As yet, the subcellular localisation of the N-terminal region of TamB remains unknown. Since to date no topological description of TamB has been reported it is currently not known whether the majority of the protein resides in the periplasm or the cytoplasm. Sequence-based analyses of the DUF490 domain from TamB indicate that the C-terminal-most 90 amino acids are separated from the rest of the domain by an α-helical bundle with one of these helices serving as a putative transmembrane helix. This would suggest that at least the last 80 amino acids may be localised to the periplasm. Therefore, an expression construct was produced containing the C-terminal fragment $TamB_{DUF490(1180-1259)}$, which consists of the final 80 amino acids of TamB and is referred to as $TamB_{C80}$ hereafter. $TamB_{C80}$ was expressed and purified from inclusion bodies using the same protocol as that for $TamB_{DUF490}$ (Figure 4-6).

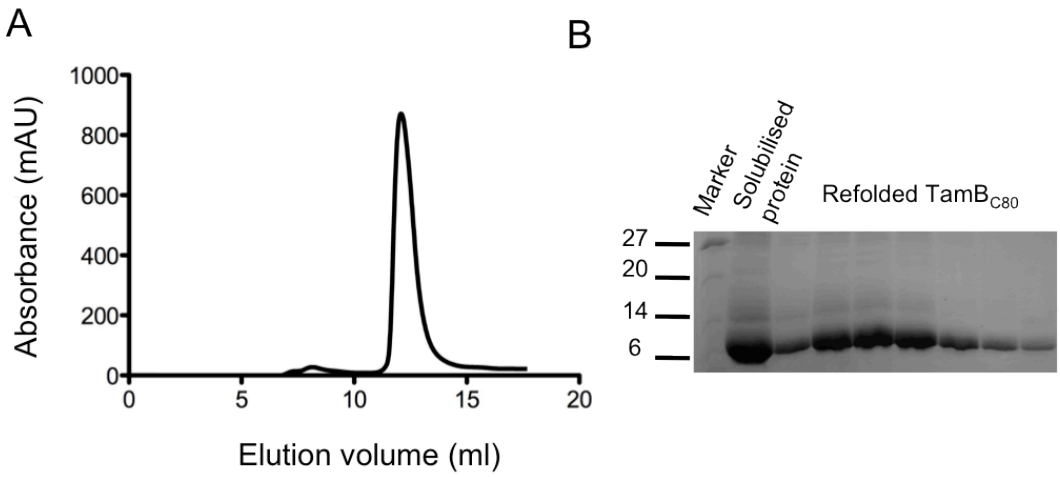


Figure 4-6 SEC and purity of $TamB_{C80}$ after refolding from IBs. (A) SEC profile of purified $TamB_{C80}$ run on a Superdex S75 analytical column. The elution volume is approximately 13 ml, indicating that the purified fragment is folded and compact. The SEC run was performed in 50 mM Tris-HCl, 50 mM NaCl, 0.4% CHAPS, which has a very small micelle size of ca 6-8 kDa. (B) 18% SDS-PAGE gel showing purified, refolded $TamB_{C80}$. The migrating M_w of the fragment is approximately 10-12 kDa, and the estimated M_w based on its amino acid sequence is 12 kDa. Marker M_w are given in kDa.

The refolded protein elutes as a close to symmetrical peak between 12-13 ml on an analytical Superdex S75 SEC column in a buffer containing 0.4% CHAPS detergent (Figure 4-6 A). The elution profile of $TamB_{C80}$ suggests that the protein exists as a monodisperse particle. To determine if the protein was refolded successfully, purified $TamB_{C80}$ was subjected to secondary structure and protein stability measurement using far-UV CD spectroscopy. Analysis of the far-UV spectrum of $TamB_{C80}$ using Dichroweb suggested a polypeptide composed of 26% α-helix and 23% β-strand with approximately 50% of random-coil (Figure 4-7), which was similar to the computed secondary structure prediction of the $TamB_{C80}$ fragment (21% α-helix, and 40% β-strand and 39% random-coil), indicating that the measured values are close to the predicted secondary structure calculations (Figure 4-3) (149). Thermal melting of $TamB_{C80}$ monitored by changes in the molar ellipticity at 218 nm shows a lack of cooperative transition which suggests that $TamB_{C80}$ lacks a well-defined tertiary structure (Figure 4-8).

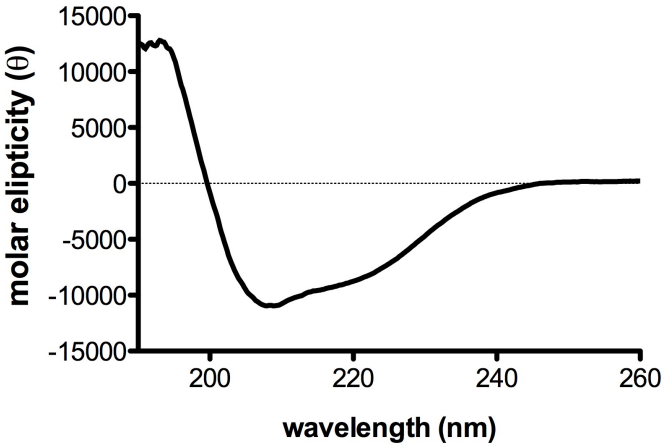


Figure 4-7 Far-UV CD spectrum of $TamB_{C80}$ indicates a polypeptide with a mixed α/β composition. Spectra were acquired for a 42 μM sample of $TamB_{C80}$ in 100 mM sodium phosphate, pH 7.5, 0.05% LDAO.

Attempts to further structurally characterise $TamB_{C80}$ were hampered by the molecule's dependence on the presence of detergent in the buffer. In the absence of any detergent $TamB_{C80}$ aggregated, as was revealed by SAXS studies (data not shown). Crystallisation trials of $TamB_{C80}$ were attempted in the presence of LDAO and CHAPS since the molecule

was shown to be stable in these detergents, however after screening 567 crystallisation conditions no successful hits were identified.

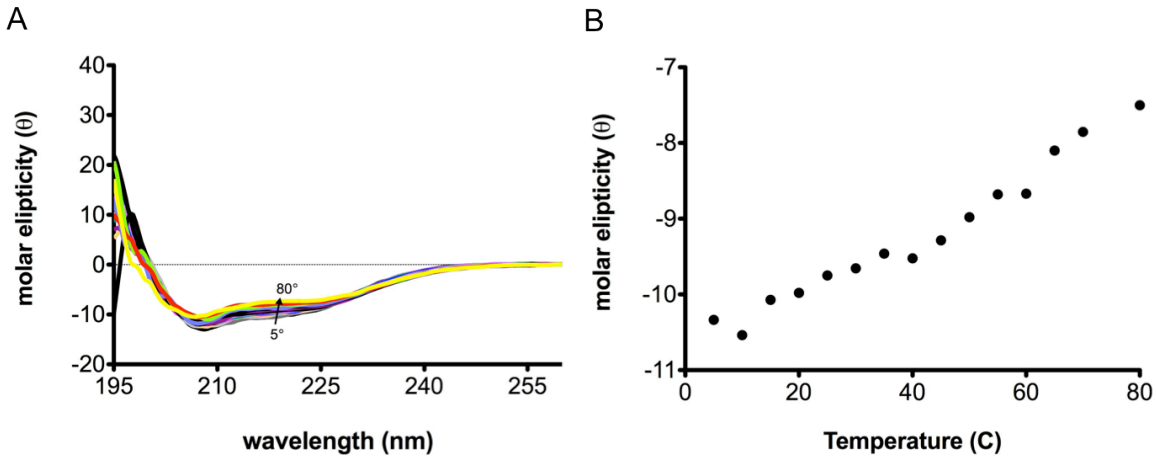


Figure 4-8 Thermal melting of $TamB_{C80}$ reveals a polypeptide molecule lacking a well-defined tertiary structure. (A) Overlay of the far-UV curves with increasing temperature shows that even at 80°C considerable secondary structure remains. Colours represent different spectra at different temperatures. (B) Plot of the loss of molar ellipticity at 218 nm with increasing temprature shows a linear decrease in molar ellipticity, preventing the determination of an accurate T_m.

4.2.4 *Expression, purification and characterisation of $TamB_{DUF490}$ N-terminal region*

The polypeptide region N-terminal of the predicted TM helix of DUF490 (DUF_{trunc1}, residues 926-1163) was overexpressed and purified. The construct was soluble in the absence of any detergents, albeit prone to slight aggregation. The elution profile of the protein showed the presence of several peaks on an analytical Superdex S200 SEC column. The leading peak corresponds to high molecular weight aggregates (>2 MDa) with several small shoulders in several peaks around 11 and 12.5 ml and a major peak at 14.6 ml (Figure 4-9). The elution volume of the major peak, according to the column calibration, meant that the protein could elute as a 60 kDa molecule (Mw calculated from the amino acid sequence is 26 kDa), suggesting that the protein may be dimeric.

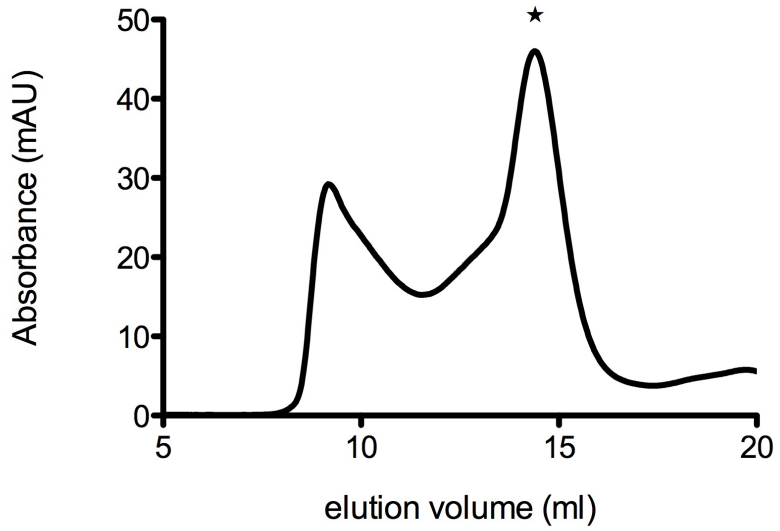


Figure 4-9 SEC profile of DUF_{trunc} suggests a soluble construct, albeit with an earlier than expected elution volume. * represents the peak used for further biophysical studies. Buffer used for SEC was 50 mM Tris-HCl, 200 mM NaCl, pH 7.5, 3% glycerol. The SEC column was an analytical Superdex S200 GL 10/300.

The complex elution profile of DUF_{trunc} suggests that the protein is prone to non-specific oligomerisation that could arise during the purification of the molecule. Purified DUF_{trunc} protein (* peak was selected, Figure 4-9) was further characterised using SAXS with scattering data acquired for a range of protein concentrations (0.4-3.2 mg ml^{-1}). No concentration-dependent variation in R_g was observed suggesting no oligomerisation/aggregation was occurring with increasing protein concentration from the selected peak (peak *). Additionally, no inter-particle interference was present (usually manifested by non-linearity of the Guinier region). The measured R_g of DUF_{trunc} was 38 ± 3 Å and the extrapolated I(0) gave a value equivalent to an M_w of 30.9 kDa, implying that the protein is monomeric in solution (Figure 4-10). Analysis of particle flexibility suggested that the molecule exhibits a high degree of flexibility as judged by the tailing effect in the Kratky plot (Figure 4-11 A). This was further emphasised by the Porod-Debye plot, which did not obey hyperbolic scattering decay for the particle (Figure 4-11 B). Instead the scattering decayed as a function of q^{-3} as observed in the $I(q)q^3$ versus q^3 plot suggesting that some inherent flexibility is present in the protein. This finding

therefore implies that DUF_{trunc} must possess several flexible/unstructured regions or long loops that give rise to the observed structural fluidity.

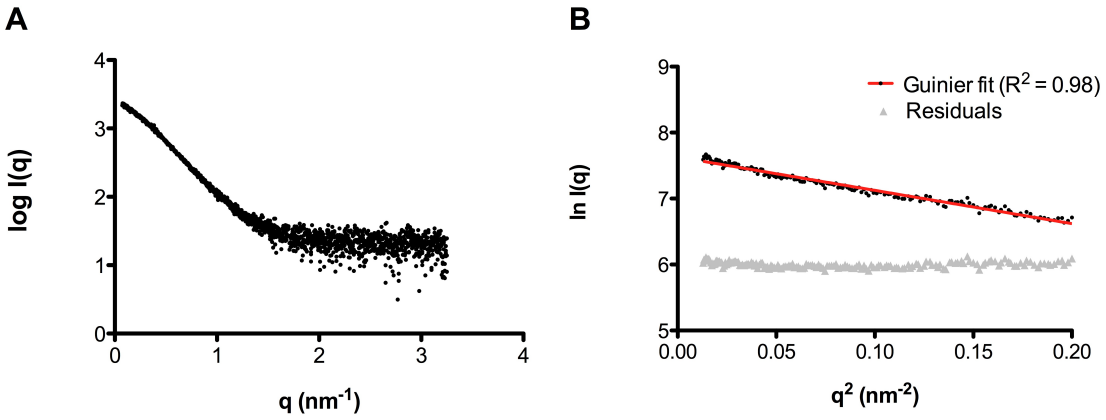


Figure 4-10 SAXS characterisation of DUF_{trunc}. (A) Raw scattering curve of DUF_{trunc} used for data analysis. 8 data points were removed from the low-angle region region of the curve due to beam-stop interference and 200 points were removed from the high-angle region due to poor data quality. (B) The linear Guinier region of DUF_{trunc} shown suggests that no detectable aggregation is present in the sample.

The p(r) plot reports on the sum of all inter-electron distances within the scattered particle and therefore provides information about the overall particle shape and maximum dimensions in solution.

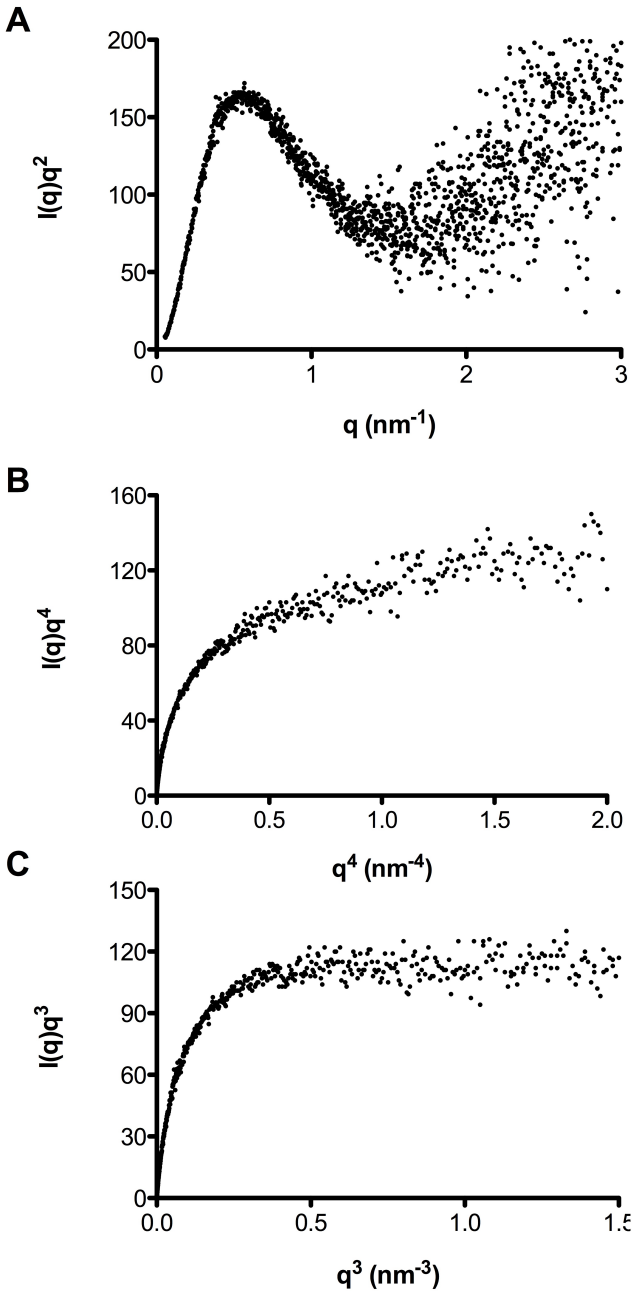


Figure 4-11 Flexibility analysis of DUF_{trunc} implies a flexible, dynamic particle. (A) The Kratky plot suggests that the particle exhibits a significant degree of flexibility in solution without any evidence for flexible multi-domain character. (B) The Porod-Debye plot depicts a clear lack of plateau usually observed for structurally homogeneous particles in solution. (C) The Kratky-Debye plot represents the scattering decay as a transformation of scattering intensity as a function of q^3 reveals a hyperbolic uniform curve suggesting that DUF_{trunc1} behaves as a highly dynamic particle.

The shape of the p(r) plot derived from the scattering data for DUF_{trunc} revealed a non-Gaussian distribution of inter-electronic distances, which extended to approximately 150 Å (D_{max}). *Ab initio* molecular envelopes were built using DAMMIF, which revealed a highly elongated particle.

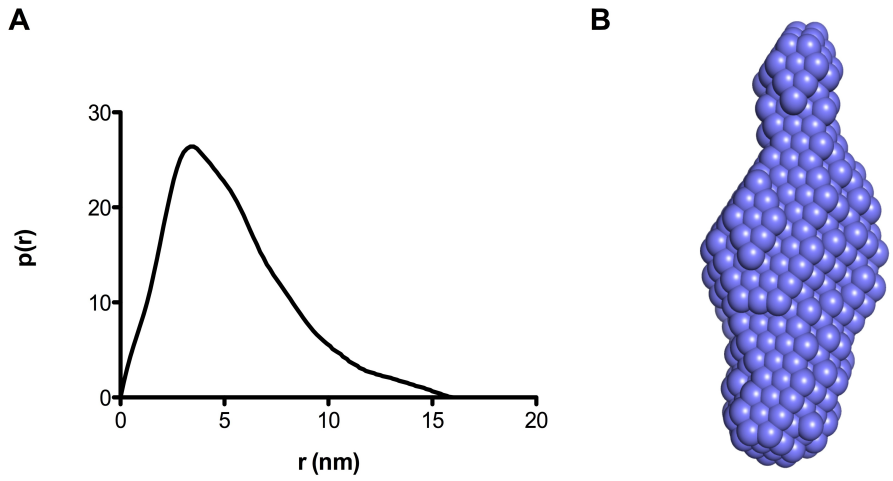


Figure 4-12 Pair distance distribution and *ab initio* shape reconstruction of DUF_{trunc} reveal an extended molecular shape. (A) The distance distribution curve shows that the molecule is highly asymmetric in solution with inter-electron distances ranging up to 15 nm, which represents the maximum length of the particle (D_{max}). (B) *Ab initio* model reconstructed from the pair distance distribution data using DAMMIF.

20 individual models were generated and averaged using DAMAVER (153). No significant deviations were observed upon visual inspection of the individual models (data not shown). The averaged models were filtered using DAMFILT in order to remove low occupancy and loosely connected atoms (153). The fit of the final *ab initio* model to the experimental scattering data is in excellent agreement, signifying that it is highly representative of the average solution envelope of DUF_{trunc}. The particle shape approximates to a prolate elipsoid with calculated dimensions of 159 x 55 Å.

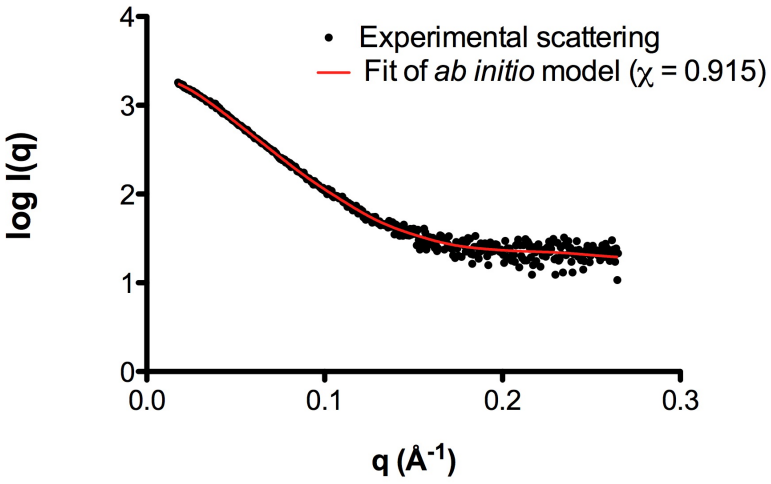


Figure 4-13 *Ab initio* model of DUF_{trunc} fits the experimental scattering data. The red line shows the calculated scattering curve represented by the *ab initio* DAMMIF model overlaid with the experimental scattering curve (black). The experimental scattering curve shown has been desmeared and smoothened during the indirect Fourier transform by the program GNOM to remove outlying scattering data points.

4.2.5 *Crystallisation of the N-terminal fragment of DUF490*

Since monodisperse DUF_{trunc} could be readily isolated and its low-resolution solution structure obtained DUF_{trunc} crystalliation trials were undertaken using the sparse-matrix crystallisation method which consisted of 384 conditions. A successful hit in a single condition (30% PEG400, 0.2 M Li_2SO_4 and 0.1 M Tris pH 8.0) was obtained which produced very thin needle-like clusters of crystals forming after several days (Figure 4-14 A). These crystals produced very poor diffraction on the in-house rotating anode and using synchrotron radiation, suggesting poorly formed contacts between the protein molecules constituting the crystal lattice. Only weak diffraction at very low resolution was observed, which confirmed that the crystals were proteinaceous in nature, exemplified in Figure 4-14 B. Numerous attempts were made to improve the quality of the crystals, such as the use of additive screen to improve packing within the crystal, as well as microseeding to improve crystal morphology and nucleation (183). *In situ* proteolysis was also attempted by supplementing crystal drops with minute quantities of trypsin and chymotrypsin (184). However, no improvement in crystal quality was observed.

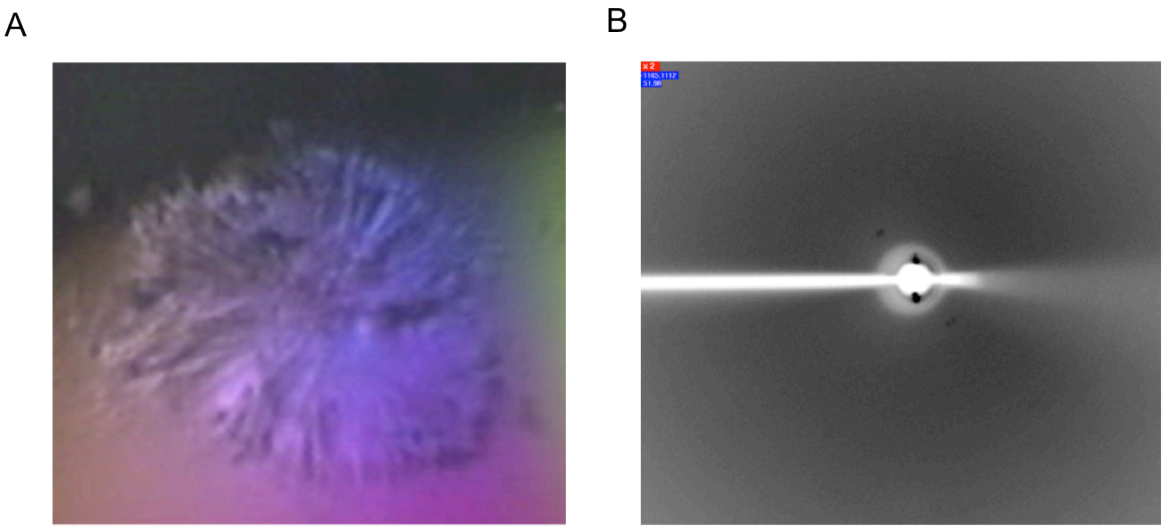


Figure 4-14 Crystals of DUF_{trunc} suggest a poorly formed lattice. (A) Thin needle-like crystal morphology of DUF_{trunc}. (B) Weak diffraction spots observed suggest that the crystals are not salt-like in nature and most likely consist of poorly packed protein molecules.

In order to try to improve the quality of the crystals, the DUF_{trunc} construct was optimised further by removing several residues from the N- and C-termini, as well as switching the His_6 tag from the N-terminus to the C-terminus of the construct, from here on referred to as $DUF_{963\text{-}1138}$. The new construct could readily be overexpressed in much higher quantities than DUF_{trunc}. SEC traces of $DUF_{963\text{-}1138}$ indicated the presence of higher M_w species (Figure 4-15). Only the peak corresponding to the last eluting species was pooled for crystallisation. Foldedness of the new construct was confirmed by far-UV CD spectroscopy and was in excellent agreement with bioinformatic prediction of the secondary structure composition (Figure 4-16).

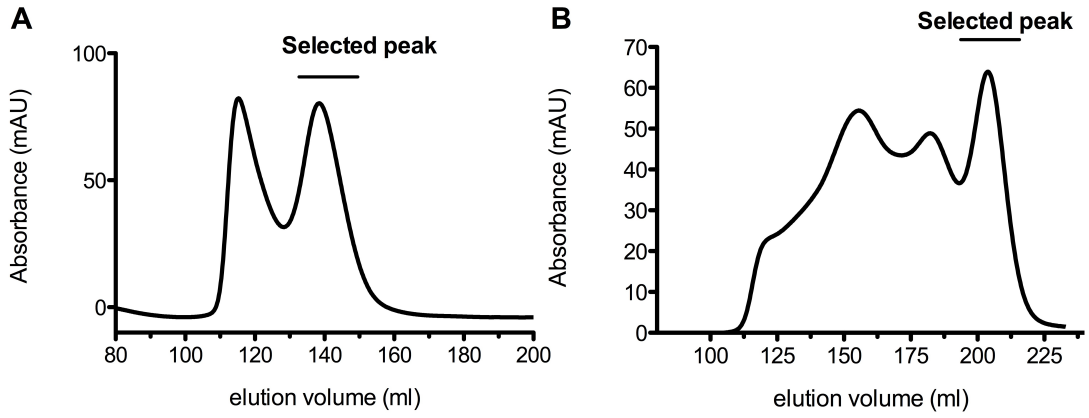


Figure 4-15 Size-exclusion chromatograms of DUF490$_{963-1138}$ from two independent purifications. Chromatograms from SEC using (A) Superdex S75 and (B) Superdex S200 are shown. Both chromatograms show the multimeric nature of the protein. For the purposes of crystallisation and CD analyses only the putative monomeric species was utilised.

DUF490$_{963-1138}$ was concentrated to 15 mg ml^{-1} and crystallisation trials were repeated as described for DUF$_{trunc}$. Barrel-shaped crystals appeared after several days in 0.1 M HEPES, 15% (v/v) PEG400, 0.2 M $CaCl_2$, pH 7.0, at 289K. After optimisation, diffraction-quality crystals were obtained in 0.1 M HEPES, 25% (v/v) PEG400 0.2 M $CaCl_2$, pH 8.0, using 8 mg ml^{-1} protein and a protein:reservoir ratio of 1:1. The crystals were cryoprotected using 0.1 M HEPES, 30% (v/v) PEG400, 0.2 M $CaCl_2$, pH 8.0, by transferring the crystals using a litho loop into the cryo solution for 3 seconds and flash-cooling in a nitrogen-gas stream at 110 K (Figure 4-17).

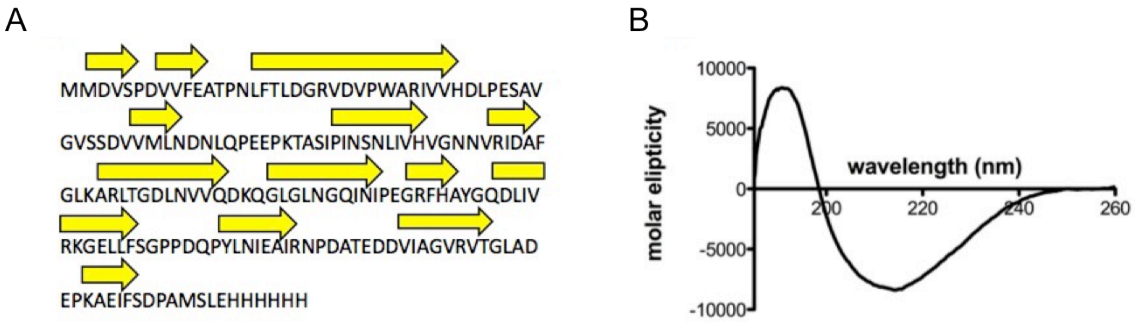


Figure 4-16 Secondary structure prediction and far-UV CD spectrum of $DUF_{963\text{-}1138}$. (A) Secondary structure prediction of the $DUF490_{963\text{-}1138}$ fold using PSI-PRED suggests that the protein consists predominantly of β-strands and random coil. (B) Far-UV CD spectrum of the construct confirms that the protein consist predominantly of β-strands (30%), turns (25%) and disordered (33%) polypeptide with minor α-helical content (12%). Values are calculated based on the average of all matching solutions. Predicted secondary structure estimates shown in (A) are 38% β-strand and 62% random coil.

Several crystals were washed in 35% PEG400 solution several times, dissolved and analysed by SDS-PAGE to ensure that the crystals consisted of the purified $DUF_{963\text{-}1138}$ construct and not a contaminating impurity (Figure 4-18).

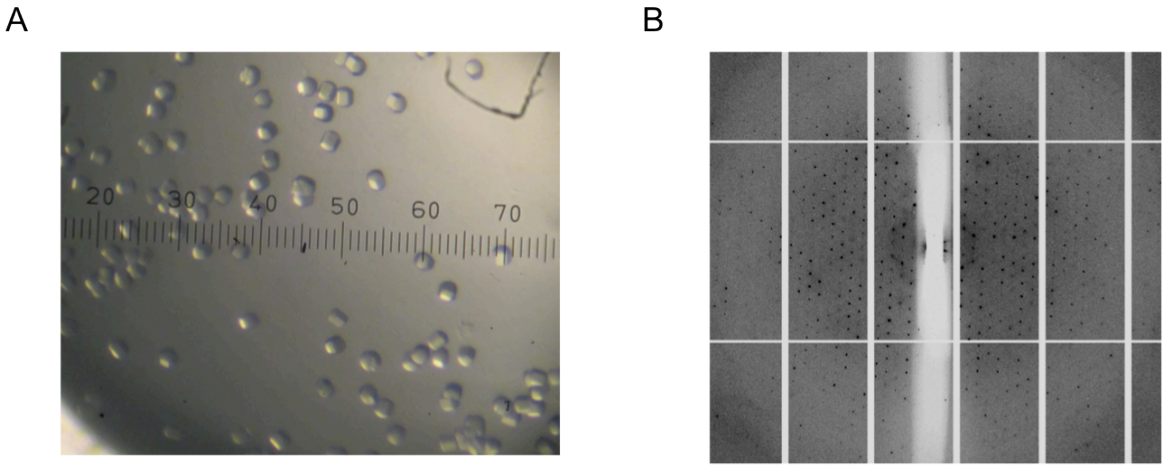


Figure 4-17 Crystals of $DUF_{963\text{-}1138}$ have a different morphology from that of the previously obtained DUF_{trunc} crystals. (A) Crystals grew within 48 hours and were no greater than 50 μm in size. (B) Representative diffraction pattern from the crystals shown in (A). Diffraction did not extend beyond 3-4 Å.

X-ray diffraction data were collected on the I03 beamline at the Diamond Light Source synchrotron (Harwell, UK). Data were collected using an oscillation angle of 0.15° with a 0.08 second exposure time. A total of 3600 frames were collected and indexed using *i*MOSFLM (161), scaled and merged using AIMLESS from the CCP4 program suite (160,162). The complete dataset could be processed to 2.42 Å with a mean *I*/sigma (*I*) *of* 5.7 for the outer shell with very high multiplicity (29.5). Two monomers are predicted in the asymmetric unit with a calculated Matthews coefficient of 2.90 $Å^3$ Da^{-1} and a solvent content of 57.6%. Analysis of possible twinning was carried out using *phenix.xtriage* and no twinning was detected (185).

Table 7 Diffraction data of a complete dataset of $DUF_{963-1138}$

Diffraction source	DLS I03
Wavelength (Å)	0.9763
Temperature (K)	100
Detector	Pilatus3 6M
Exposure time (s)	0.08
Rotation per image (˚)	0.15
Space group	$P3_121$
Cell dimensions, *a, b, c* (Å)	57.34, 57.34, 220.74
Resolution (Å)	48.45-2.42 (2.48-2.42)
No. of unique observations	16922 (1222)
Multiplicity	27.6 (29.5)
Completeness (%)	99.7 (98.9)
R_{merge} (%)	13.7 (101.0)
R_{pim} (%)[a]	2.7 (20.7)
Mean *I*/sigma (*I*)	21.4 (5.7)

[1] Values in parentheses refer to the highest resolution shell

[a] $R_{pim} = \Sigma_{hkl}[1/(N - 1)]^{1/2}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i I_i(hkl)$

Due to the unavailability of any structural homologues of DUF490, *ab initio* structural homology modelling was attempted using PHYRE, iTASSER and SWISS-MODEL. However, the confidence in the models was very poor and molecular replacement was unsuccessful with any of these models (181,182,186). Therefore, direct methods must be applied in order to solve the phase problem and obtain the structure of $DUF_{963-1138}$. Attempts to obtain experimental phases with heavy metal derivatisaton as well as selenomethionine-labelled protein for single-wavelength anomalous dispersion (SAD) phasing were

attempted but at the present time have not yielded a structural model.

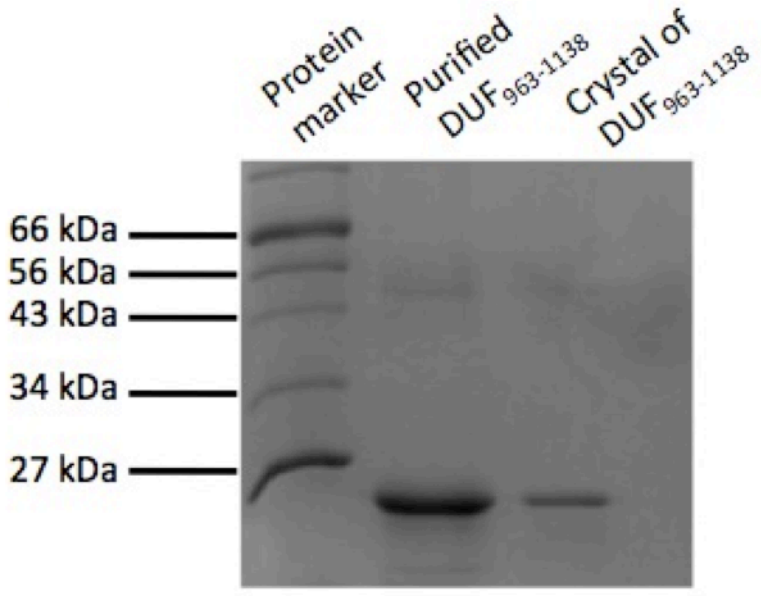


Figure 4-18 15% SDS-PAGE of purified $DUF490_{963-1138}$ shows a single band migrating at the approximate molecular weight expected for the construct (20 kDa). Additionally, several crystals were washed with 35% PEG400 solution three times, prior to melting in pure H_2O and loaded onto the gel.

4.3 Discussion

In this chapter, refolding and purification of the full-length DUF490 domain is presented, along with purification and characterisation of two major fragments of DUF490, the C-terminal 80 amino acids (residues 1180-1259) as well as the N-terminal constructs corresponding to residues 926-1163 and 963-1138. Previous studies implicated the last 80 amino acids in the interaction with the POTRA domains from TamA (138). Far-UV CD experiments as well as computational secondary structure predictions of the 80 C-terminal amino acids presented in this chapter revealed a β-strand rich polypeptide chain suggesting that TamA-TamB complex formation may rely on β-augmentation as a mechanism of association.

The N-terminal region of DUF490 exhibited a propensity for oligomerisation during purification and this is likely to be an artefact of the purification process rather than an indication that the protein functions as an oligomer. Previous work has shown that TamB behaves as a monomer on blue-native PAGE (138). Similarly the presence of higher oligomeric species was observed during the purification of BamB and BamE as reported by Albrecht and Zeth (72). Since the predominant secondary structure element of N-terminal DUF490 polypeptides is the β-strand this could explain the non-physiological oligomerisation of DUF490 *in vitro* through, for example, β-strand swapping as seen in the crystal structure of C-cadherin, or through triangular β-sheet prism formation seen in the structure of the P22 tailspike (187,188). Interestingly, TMpred analysis of the DUF490 sequence predicts a putative transmembrane helix within that region of the domain, with an inside-to-outside topology.

TamB has previously been shown to possess an N-terminal signal sequence α-helical motif that anchors the protein to the IM (138). The predicted topology of this N-terminal helix, according to TMpred, is outside-to-inside, which would place the majority of TamB inside the cytosol, with the potential for the C-terminal region (i.e. the last 80 amino acids) to be periplasmic. However, early studies of IM topologies have suggested a "positive-inside" rule, implying that the N-terminal positively charged residues are likely to remain cytoplasmic during the insertion of the helical motif into the membrane (189). TamB has two Lys residues within the first 6 amino acids before the start of the transmembrane helix (predicted residues 7-26), which might suggest that the N-terminus of TamB is cytoplasmic, thus placing the rest of the protein in the periplasm. Further biochemical experiments are needed in order to discern the cellular topology of the whole TamB protein and the location of the DUF490 domain.

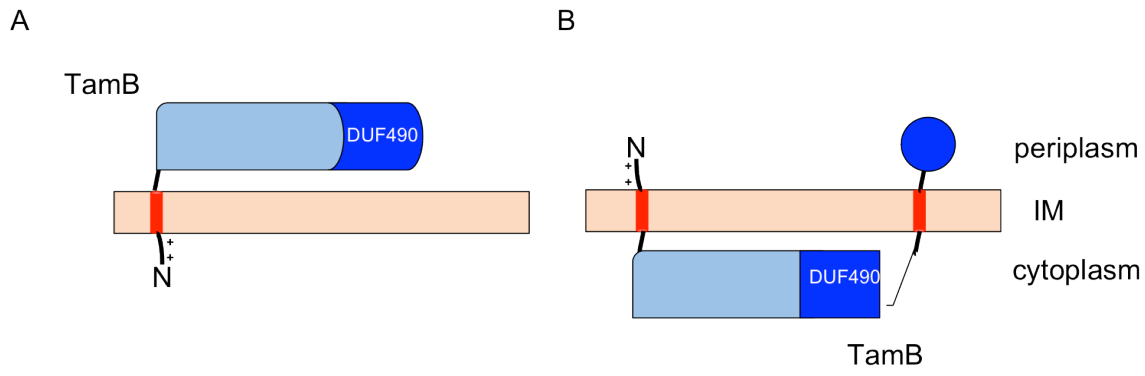


Figure 4-19 Predicted models of the membrane organisation of TamB protein. (A) The model relying on the "positive inside" rule suggests that positively charged residues prior to the transmembrane helix (red) are located in the cytoplasmic region of the cells, thus putting the rest of the TamB molecule into the periplasm. (B) Topological arrangement of TamB as suggested by TMpred.

The structural characterisation of the N-terminal region of DUF490 domain using SAXS has revealed that it forms an extended particle with a high degree of flexibility. This flexibility and dynamic nature of the protein could be the cause of poor crystalline diffraction of the DUF_{trunc} crystals, due to the presence of unstructured regions or numerous loops. Truncation of the N-terminal domain to a more compact molecule led to the formation of better diffracting crystals; however, phase information is required in order to calculate the electron density for model building. Experimental phase determination is currently in progress. Determination of the high-resolution structure of a part of DUF490 may help to define the functional role of this conserved domain. The C-terminal region of DUF490, on the other hand, possesses secondary structure elements, yet appears to lack a defined tertiary structure as judged by the thermal unfolding experiments. Therefore, the DUF490 domain may consist of an N-terminal extended, prolate structure rich in β-sheets, separated from the C-terminal 80 amino acids, also rich in β-sheets, by an α-helical bundle.

4.4 Conclusion

The C-terminal DUF490 domain from TamB was successfully refolded from inclusion bodies and purified to homogeneity. The protein was stable in the presence of Triton X-100 detergent and eluted abnormally early during SEC, which was attributed to the anisotropic shape of the protein. Further structural characterisation of the N-terminal region of DUF490 (residues 926-1163), using SAXS revealed that the particle exists as a folded, extended shape in solution with a high degree of flexibility and poor propensity for ordered crystalline packing. Improvement in the lattice packing was achieved by the truncation of the N-terminal construct and well-diffracting crystals were obtained. The C-terminal 80 amino acids of DUF490 were shown to possess regular β-sheet rich character, without any putative tertiary fold and are thought to be separated from the rest of the DUF490 domain by an α-helical bundle as predicted by TMpred and Jpred (176,189).

5 Biophysical characterisation of TAM complex formation

Circular dichroism data presented in this chapter has been performed in collaboration with Dr Sharon Kelly

All NMR spectroscopy was performed under the supervision of Dr Brian O Smith

5.1 Introduction

TamA and TamB have previously been shown to form a periplasm-spanning complex in Gram-negative bacteria (138). Initial characterisation of the complex had revealed that the N-terminal periplasmic POTRA domains of TamA and the C-terminus of the DUF490 domain from TamB were the main binding determinants within the complex (138). The main objective of this study was to further characterise the interactions between TamA and TamB using several biophysical techniques in an attempt to shed light on the mechanism wherein the TAM complex is formed, thus gaining insight into the roles of POTRA domains as protein platforms of macromolecular assemblies. Current mechanistic insight into the intermolecular association involving POTRA domains comes primarily from X-ray crystallographic studies of these domains. The first atomic models of POTRA1-4 from BamA had suggested that one possible means of substrate chaperoning of these periplasmic domains must involve β-augmentation, a mode of binding wherein a β-strand from a neighbouring molecule forms an extended intermolecular β-sheet with another molecule (Figure 5-1).

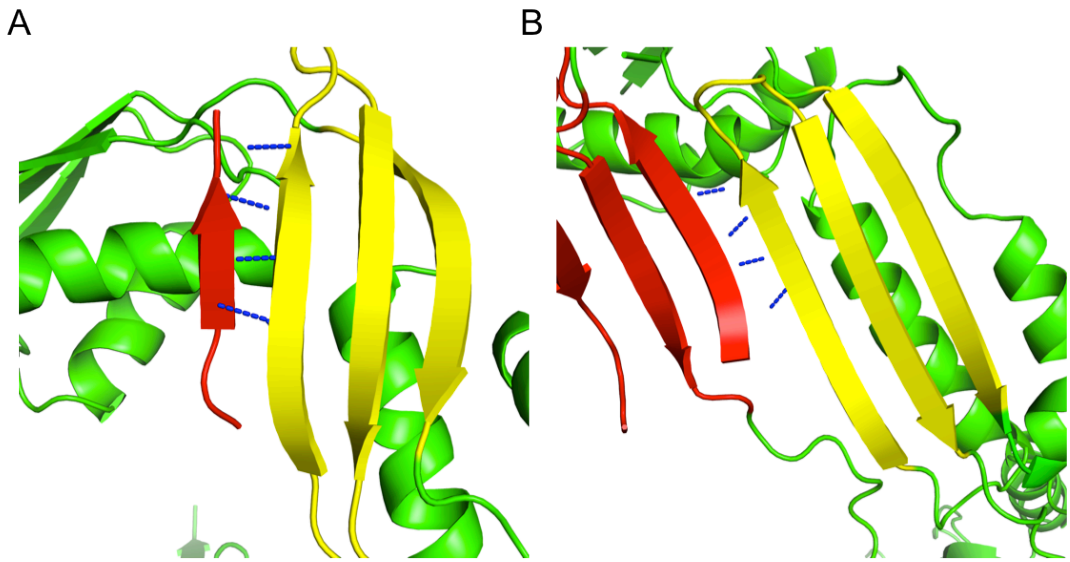


Figure 5-1. Evidence of β-augmentation revealed by molecular packing of BamA POTRA1-4 (A) and TamA POTRA1-3 (B) PDB IDs: 2QCZ and 4BZA respectively. In yellow, the β-sheets from one POTRA domain augment the β-sheets/strand from the adjacent domain (red) of a symmetry-related molecule in the crystal lattice (30,140). Hydrogen bond contacts between the augmenting β-strands are shown as blue dashes.

All the components of the BAM complex reside in the bacterial periplasm where they behave as lipoproteins anchored to the OM. Several studies have addressed protein-protein interactions involving POTRA domains and their partner proteins, however to date there is no clear structural or biophysical description of these complexes. Despite the availability of high-resolution X-ray structures of all the components of the BAM complex, there is little structural and biophysical description of the complex formation. BamA POTRA domains have been shown to directly interact with BamB through genetic and cross-linking studies (69). Docking studies have implicated POTRA domains 2-5 to be involved in BamA-BamB complex formation, with POTRA3 being the main binding determinant (30,66). In-depth analysis of the crystallographic model of POTRA3 revealed the presence of an additional β-bulge, which is absent from other POTRA domains. Deletion of this bulge abolished interactions with BamB suggesting that subtle secondary structure elements present in POTRA domains could play a major role in the formation of POTRA-mediated macromolecular assemblies (30). Further studies of lethal mutants and domain deletions of BamA in *E. coli* and *N. meningitidis* have revealed that POTRA5 is involved in complex formation with BamD (98,190).

Extensive biochemical and biophysical characterisation of the interactions between the POTRA domains of FhaC, part of the TPSS, and its sole substrate FHA have revealed that the binding affinity is in the low micromolar range (94). The affinity of FhaC POTRA domains markedly differs from the reported binding interactions between the BamA POTRA domains and hydrophobic peptides derived from a BamA substrate molecule, OMP PhoE. Electrophysiology experiments have also shown that unfolded OMP substrate porins could increase the channel size of Omp85 from *N. meningitides* through the recognition of the C-terminal signature motifs (Aro-X-Aro residues) present in all OMPs (56). Closer inspection of the amino acid sequence of TamB DUF490 revealed the presence of a motif at the extreme C-terminus (YQFEF) that mimics the BamA recognition motif found in OMPs that consists of a repeat of two aromatic residues separated by another residue, which is necessary for the opening of the BamA β-barrel and membrane incorporation of substrate proteins (56). Moreover, a similar motif has been shown to be involved in inducing the envelope stress response by interacting with an IM protease DegS (191). Since there is no experimental evidence of TamB incorporation into the OM,

this observation leads to two hypotheses about TamB: (a) TamA-TamB interactions occur via the C-terminal motif of TamB in a manner similar to the β-augmentation envisaged in the BamA-BamB complex; (b) there exists a possibility of BamA-TamB complex formation at some stage of the AT biogenesis process.

The aims of this chapter address the former hypothesis to delineate the binding interfaces of TamA and TamB as well as provide a detailed account of TamA-TamB interactions.

5.2 Results

5.2.1 *Biophysical characterisation of TamA-TamB interactions*

Structural characterisation of refolded TamB DUF490 domain was described in Chapter 4. In order to further validate the refolding of $TamB_{DUF490}$, due to the unavailability of a functional assay, preliminary $TamB_{DUF490}$-$TamA_{POTRA}$ interactions served as the refolding quality control. Complex formation, previously demonstrated using analytical SEC, was assayed using analytical SEC and chemical cross-linking between the two proteins (138). Each protein was passed down a SEC column individually as a control for its elution volume, and a sum elution profile was calculated from each of the protein elution profiles alone as a hypothetical profile for non-interacting species. For the binding reaction, a 3-fold excess of $TamB_{DUF490}$ was mixed with 20 μM $TamA_{POTRA}$ and left to equilibrate at room temperature for 15 minutes. Prior to injection, samples were placed in a 37°C water bath for 2 minutes and then injected onto the column. Examination of the elution profile of the $TamB_{DUF490}$-$TamA_{POTRA}$ complex indicated a small shift in the elution volume of $TamB_{DUF490}$, as well as the presence of a leading boundary, which did not overlap with the leading front of $TamB_{DUF490}$ nor the calculated profile for both proteins in isolation, suggesting that the two proteins were undergoing a weak, reversible heterodimerisation (Figure 5-2 A). Since the sample volume involved in the reaction (0.5 ml) undergoes a large dilution once it has been injected onto the column (bed volume 24 ml, void volume 8.5 ml) it can be estimated that the K_d of the interaction is somewhere in the micromolar range. In the case of high-affinity, non-reversible complexes with a K_d in the nanomolar-

picomolar range, the complex would be expected to elute earlier than the individual components, as a separate peak.

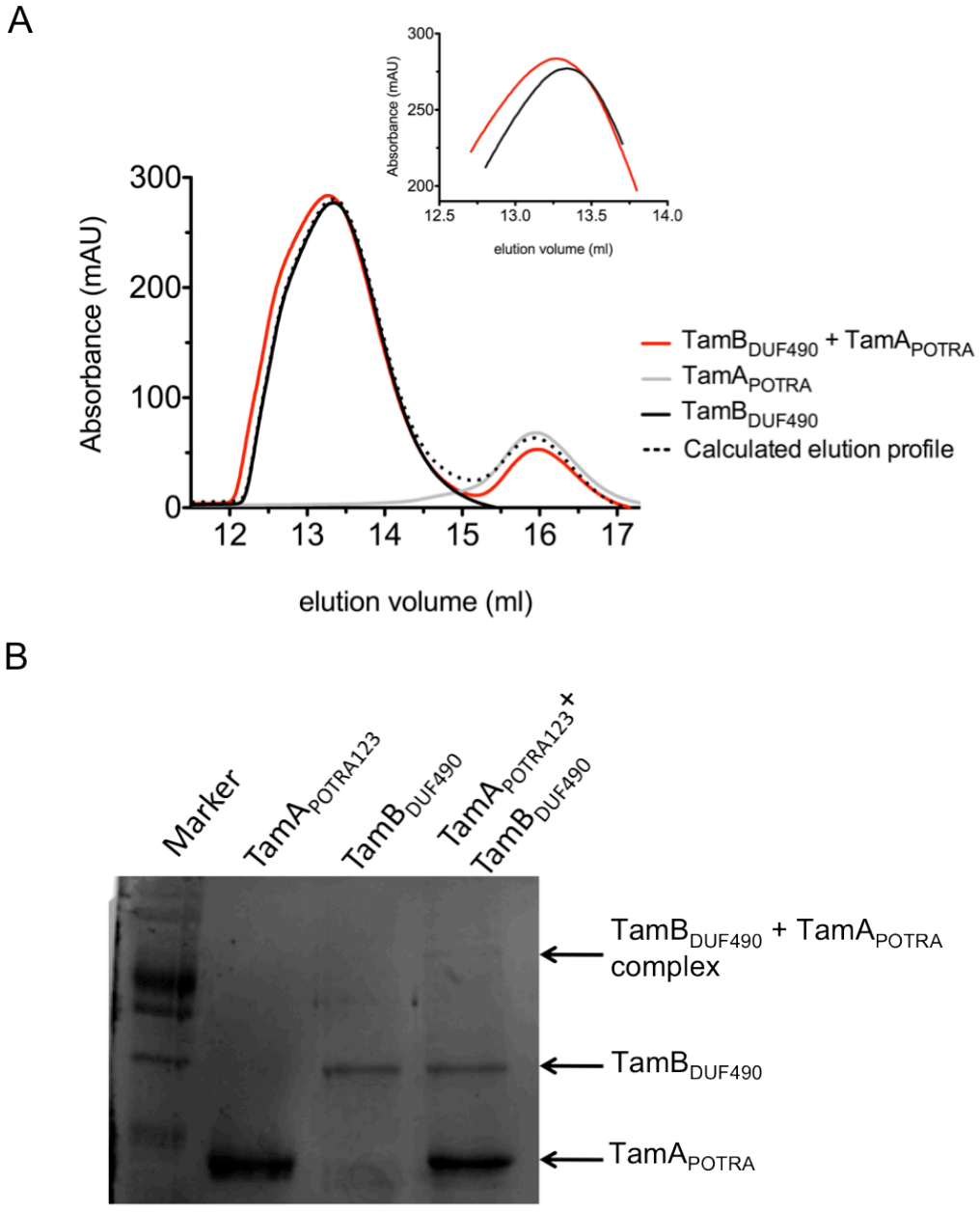


Figure 5-2 Validation of TamB$_{DUF490}$ refolding through TamA-TamB interaction studies. (A) Analytical size-exclusion chromatography. Inset shows the zoom on the peak to emphasise the change in peak position. SEC reactions were carried out in buffer consisting of 300 mM L-arginine, 500 mM NaCl, 50 mM Tris pH 7.5, 0.015% Triton X-100. (B) SDS-PAGE gel showing cross-linked product (lane 4) between TamA$_{POTRA}$ (lane 2, 20 μM) and TamB$_{DUF490}$ (lane 3, 40 μM). Estimated molecular weight of cross-linked product is ca. 70-80 kDa, individual proteins have a M_w of 28.6 kDa (TamA$_{POTRA}$) and 42 kDa (TamB$_{DUF490}$). All reactions were carried out in 100 mM sodium phosphate, pH 7.5, 300 mM NaCl, 0.015% Triton X-100. 20 μM TamA$_{POTRA}$ was mixed with 40 μM TamB$_{DUF490}$ in the presence of a 20-fold excess of DTSSP and left on ice for 1 hour.

Chemical cross-linking served as an additional control for the TamA-TamB interaction. $TamA_{POTRA}$ and $TamB_{DUF490}$ alone served as controls. Cross-linked products were analysed on a 15% SDS-PAGE gel under non-reducing conditions to maintain the cross-linker intact (Figure 5-2 B). A faint band of estimated M_w of 70-80 kDa was seen in the cross-linked sample, which was not present in the $TamA_{POTRA}$ or $TamB_{DUF490}$ samples alone, which corresponded to the cross-linked TamAB complex (the estimated M_w of the potential complex is 70.4 kDa). The presence of both proteins in the cross-linked band was later confirmed by MALDI-TOF mass spectrometry (data not shown). The above data suggest that $TamB_{DUF490}$ was refolded successfully and is able to interact with its partner protein $TamA_{POTRA}$.

Recombinant expression of $TamB_{DUF490}$ therefore paves the way for further biophysical characterisation of the TAM complex formation through quantitative techniques. The interaction between $TamB_{DUF490}$ and $TamA_{POTRA}$ was analysed further using isothermal titration calorimetry (ITC). Strong exothermic heats of binding were observed on titration of $TamB_{DUF490}$ into $TamA_{POTRA}$ (Figure 5-3). Fitting these data with a single-site binding model gave an equilibrium dissociation constant (K_d) of 5.9 ± 0.5 μM. Complex formation was enthalpically driven (ΔH is negative) and entropically disfavoured ($T\Delta S$ is negative), which is suggestive of intermolecular contact formation, involving structural rearrangement and conformational restriction of protein side-chains.

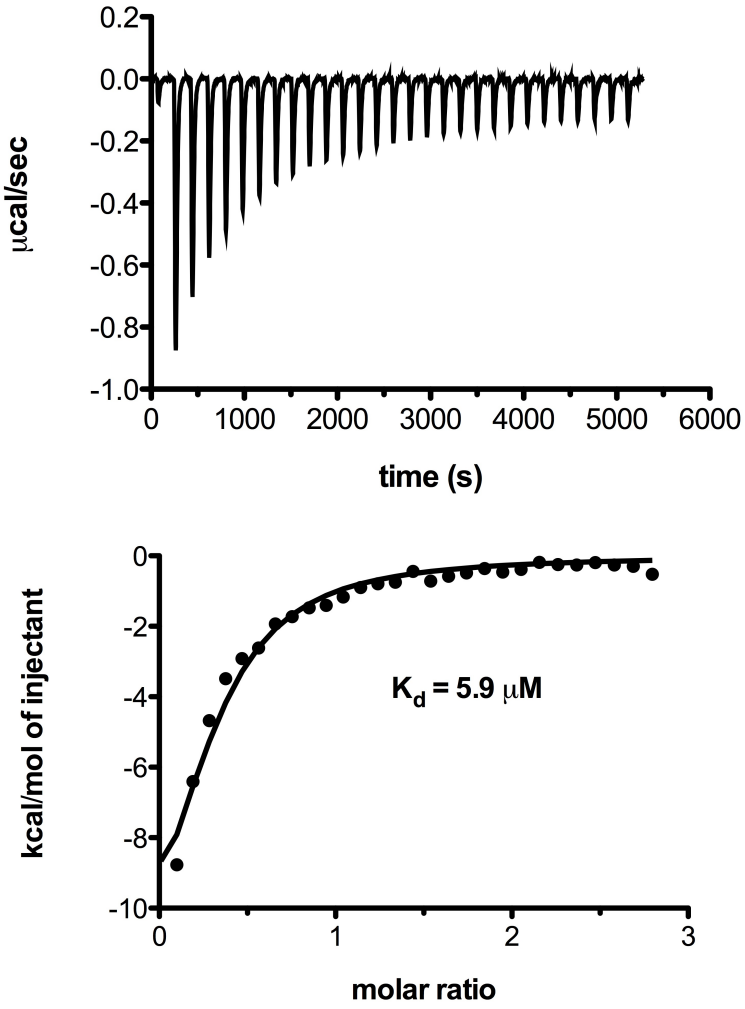


Figure 5-3 ITC analysis of the $TamA_{POTRA}$ and $TamB_{DUF490}$ interaction. Approximately 500 µM $TamB_{DUF490}$ was titrated into 20 µM $TamA_{POTRA}$. The integrated heat data were fitted with a single-site binding model**.** The reaction was performed at 25°C in 0.2 M sodium phosphate buffer, pH 7.5, 0.015% Triton X-100. Heats of dilution were obtained by titrating $TamB_{DUF490}$ into the buffer.

5.2.2 *Identification of $TamA_{POTRA}$ domains involved in TAM complex formation*

TamA possesses three tandem POTRA domains at its N-terminus, yet the contribution of each of these domains to the binding of TamB is not known. To elucidate whether all or a subset of POTRA domains are necessary for the TAM complex formation, truncations of POTRA domains were produced, deleting POTRA3 and POTRA23 from the $TamA_{POTRA123}$ construct. $TamA_{POTRA1}$ and $TamA_{POTRA12}$ could be produced recombinantly and readily purified as described in Chapter 3. Attempts to purify isolated $TamA_{POTRA2}$ and $TamA_{POTRA23}$ were unsuccessful as the proteins could not be successfully refolded after isolation from inclusion bodies. Therefore, only $TamA_{POTRA1}$ and $TamA_{POTRA12}$ were used for subsequent studies.

The effects of POTRA domain deletions on the interaction of $TamA_{POTRA}$ with $TamB_{DUF490}$ were analysed using ITC in order to try to determine which domain is involved in mediating complex formation. Titration of $TamB_{DUF490}$ into $TamA_{POTRA12}$ produced an exothermic reaction profile and the binding isotherm was fitted with a single-site model to produce a K_d of 20 μM (Figure 5-4).

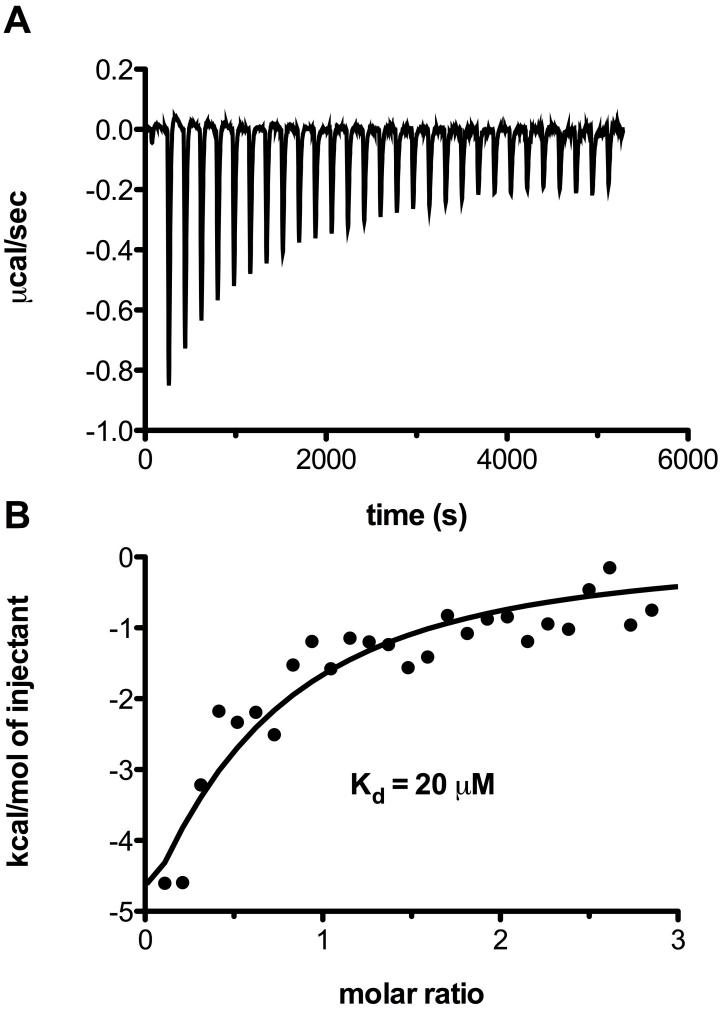


Figure 5-4 Localisation of the $TamB_{DUF490}$ binding site on $TamA_{POTRA}$. ITC was performed with $TamA_{POTRA12}$ (15-20 µM) in the cell, titrating in $TamB_{DUF490}$ (300 µM). The raw thermogram is shown (A), integrated heats and a single-site model fitted to the data is shown in (B).

The interactions between $TamA_{POTRA12}$ and $TamB_{DUF490}$ were detectable despite the molten globule-like state of the second POTRA domain within the polypeptide, which suggests that the tertiary structure the second POTRA domain may not be required for the association to take place. Moreover, it has been considered that molten-globules may possess some tertiary structure, which is not very distant from the native structure of the molecule, and may even be enzymatically active as has been demonstrated for chorismate mutase and cytochrome C (192,193). It is plausible that upon association with $TamB_{DUF490}$, the structural equilibrium of $TamA_{POTRA12}$ is shifted closer to the native state

in the POTRA2 domain as is the case with the molten-globule state of chorismate mutase which exhibited a folding transition upon binding of its transition state analogue (193).

The overall thermodynamic profile of the binding was unaltered with a favourable enthalpic contribution (ΔH is negative) and an entropic cost ($T\Delta S$ is negative). The entropic cost of the binding increased 2-fold, compared with the full-length $TamA_{POTRA}$, with a 1.4-fold increase in the enthalpy of binding. The 4-fold reduction in K_d, compared with the full-length $TamA_{POTRA}$, suggests that POTRA3 may contribute somewhat to the interactions between $TamA_{POTRA}$ and $TamB_{DUF490}$ however it does not serve as the main binding site for the molecule. Having established that $TamA_{POTRA12}$ is able to associate with TamB, interactions between $TamA_{POTRA1}$ and $TamB_{DUF490}$ were investigated by titrating 1-2 mM of $TamA_{POTRA1}$ into 40 μM of $TamB_{DUF490}$. No saturable heats of binding were observed and the titration produced the same heats as the control reaction ($TamA_{POTRA1}$ into ITC buffer) (Figure 5-5).

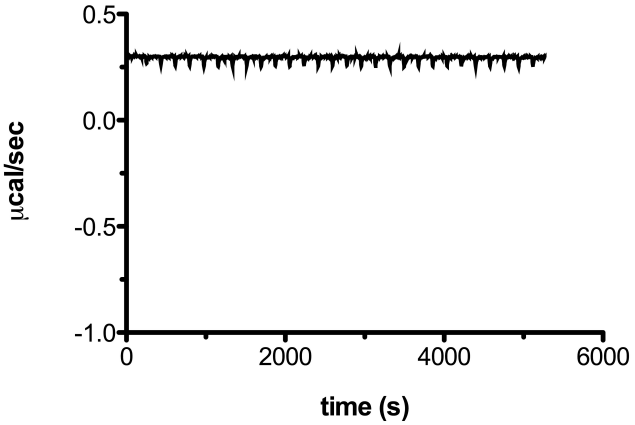


Figure 5-5 $TamA_{POTRA1}$ does not associate with TamBDUF490. Raw thermogram showing titration of $TamA_{POTRA1}$ (1-2 mM) into $TamB_{DUF490}$ (40 μM). The flat baseline indicates no heats of association during injections.

This suggests that $TamA_{POTRA1}$ is not directly involved in the association with $TamB_{DUF490}$, implying that $TamA_{POTRA2}$ serves as the predominant binding domain necessary for TAM

complex formation, with some additional contribution to the association reaction from $TamA_{POTRA3}$.

5.2.3 *Delineation of the $TamA_{POTRA}$ binding site on the $TamB_{DUF490}$ domain*

A previous study implicated the C-terminal 80 amino acids of TamB in the association with POTRA domains from TamB (138). Interactions between refolded $TamB_{C80}$ and $TamA_{POTRA}$ were analysed using ITC by titrating 2 mM $TamB_{C80}$ into 20 μM $TamA_{POTRA}$. A saturable thermogram with exothermic heats was observed and after integration of the heats, produced a binding isotherm that could be fitted with a single binding site model. The determined dissociation constant was approximately 5 μM, which is very close to the K_d determined for full-length $TamB_{DUF490}$ suggesting that the binding site within TamB for TamA is localised to the C-terminal 80 amino acids of the protein, as previously considered (138). Interestingly, the thermodynamic profile of the reaction was markedly different from the interaction with full-length $TamB_{DUF490}$. The binding reaction was primarily entropically driven with a small enthalpic contribution. The entropic drive for the association indicates that conformational re-arrangements upon binding are driving the interaction with the small enthalpic element arising from direct protein-protein contacts.

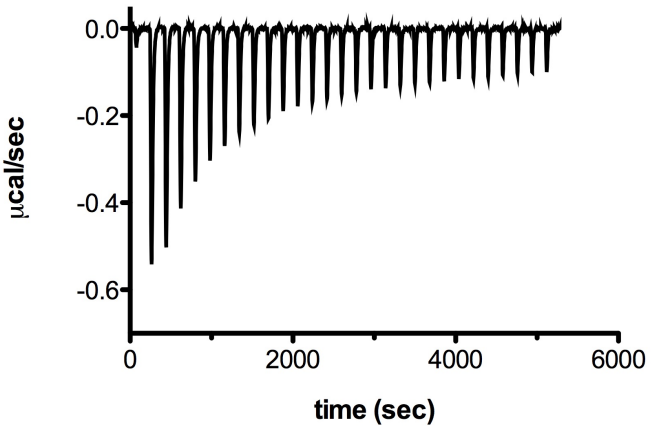


Figure 5-6 Binding isotherm for the interaction between $TamA_{POTRA}$ (20 μM) and $TamB_{C80}$ (2 mM). The K_d for the interaction between $TamA_{POTRA}$ and $TamB_{C80}$ is very similar to that determined for the $TamB_{DUF490}$-$TamA_{POTRA}$ complex which suggests that the C-terminal 80 amino acids possess all the necessary binding determinants for TAM complex formation.

Inspection of the Kyte-Doolittle hydropathy plot for the $TamB_{DUF490}$ domain, which displays the hydrophobicity of the amino acids in a protein sequence, reveals that in comparison with the rest of the DUF490 domain the last 80 amino acids possess an overall greater proportion of hydrophobic character (Figure 5-7).

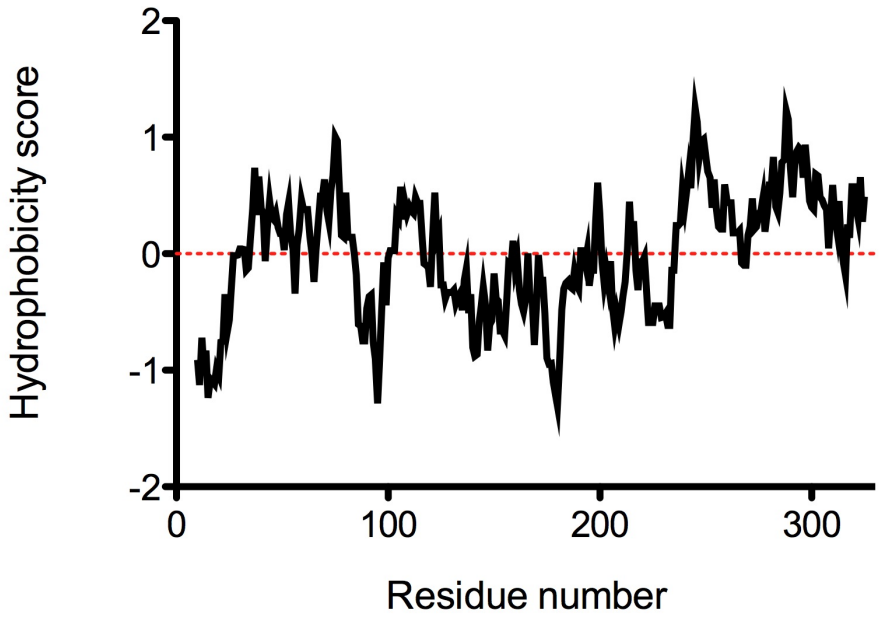


Figure 5-7 Kyte-Doolittle plot of $TamB_{DUF490}$ showing the increased hydrophobic character of the C-terminal 80 amino acids.

The increased hydrophobic character may explain the entropy-driven association between TamA and TamB that arises through desolvation effects upon association and re-ordering of the water molecules associated with the binding interfaces. The crystal structures of $TamA_{POTRA}$ domains reveal a highly hydrated interface within the POTRA2 domain, the same interface that is involved in coordinating the hydrophobic peptide *in crystallo*. Therefore, it is likely that association of TamA and TamB could lead to the re-organisation of this hydrated interface upon complex formation.

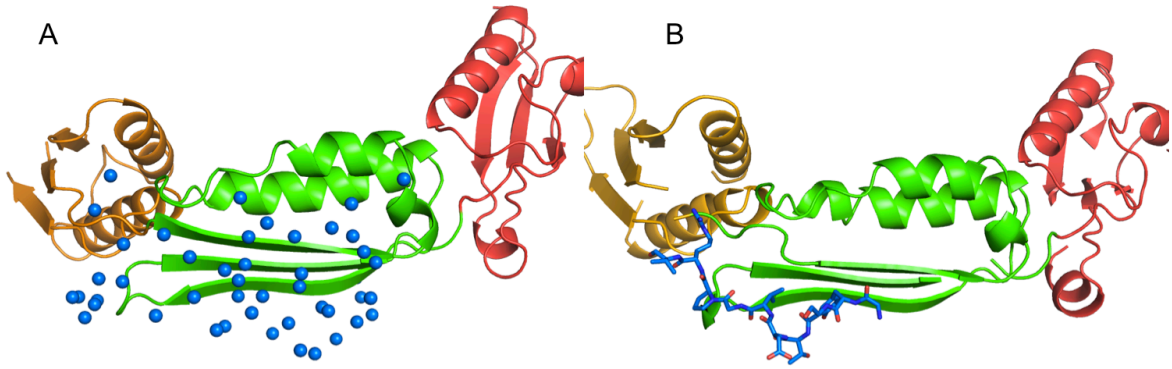


Figure 5-8 Inspection of the interface of $TamA_{POTRA(22-265)}$ shows a highly hydrated surface. (A) Highlighted water molecules around the second TamA POTRA domain (PDB ID: 4BZA) (B) Emphasis on the peptide contacts across the hydrated interface of POTRA2 which could explain the entropy-driven $TamA_{POTRA}$-$TamB_{C80}$ interactions (PDB ID: 4QAY).

Having localised the binding site to the last 80 amino acids of TamB, TamA-TamB interactions were further probed. Initially, the hypothesis was that TamB would interact with TamA via the extreme C-terminal amino acid residues, which bear a signature motif (YQFEF) found commonly in C-terminal residues of OMPs. This motif has previously been shown to increase the opening of the BamA barrel embedded in proteoliposomes, suggesting its role as a recognition motif for the BAM complex (56). Additionally, hydrophobic β-strand peptides derived from BamA substrates have been shown to interact weakly with BamA POTRA domains (58). Since far-UV melting experiments with $TamB_{C80}$ revealed that the domain does not possess any stable tertiary structure (Figure 4-8), truncations at the C-terminus of TamB were made by deleting amino acids corresponding to the last 2 and last 3 putative β-strands (20 residues and 35 residues, respectively) found in $TamB_{DUF490}$ that were predicted from the bioinformatic analyses (Figure 4-3). The constructs were refolded and purified using the same protocol as that used for $TamB_{DUF490}$. ITC was used to assess the binding capability of these $TamB_{DUF490}$ mutants for $TamA_{POTRA}$.

Deletion of the last 20 amino acids from $TamB_{DUF490}$ resulted in a moderate reduction in binding affinity towards $TamA_{POTRA}$ (K_d = 31 μM), compared with the full-length $TamB_{DUF490}$, but binding was not abolished, indicating that the last 20 amino acids do not

possess the necessary binding motif for $TamA_{POTRA}$ (Figure 5-9). However, when 35 amino acids are absent from the C-terminus of $TamB_{DUF490}$ no saturable binding was detected for $TamA_{POTRA}$ (Figure 5-9). This suggests that the amino acid segment found at the C-terminus of $TamB_{DUF490}$, although not localised to the last 20 amino acids, is involved in the interaction with $TamA_{POTRA}$.

Overall, the truncation experiments suggest that the residues at the extreme C-terminus of TamB, yet not the last 20 amino acids, are involved in the formation of the cell-envelope spanning TAM complex, mostly through the association with $TamA_{POTRA2}$.

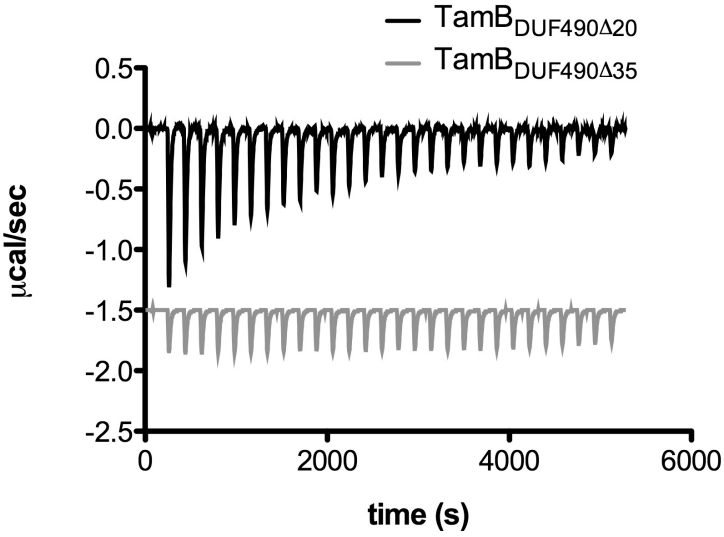


Figure 5-9 The C-terminal region of $TamB_{DUF490}$ is involved in $TamA_{POTRA}$ association. Raw ITC thermogram showing interactions between $TamA_{POTRA}$ (20 μM) and $TamB_{DUF490}$ truncation mutants, $TamB_{DUF490\Delta20}$ (300 μM) and $TamB_{DUF490\Delta35}$ (300 μM).

5.2.4 *Analysis of $TamA_{POTRA}$-$TamB_{DUF490}$ interactions using stopped-flow fluorimetry*

To further validate the interactions between $TamA_{POTRA}$ and $TamB_{DUF490}$, pre-steady state protein association kinetics were measured using stopped-flow fluorimetry under pseudo-first order conditions. $TamA_{POTRA}$ (10 μM) rapidly mixed with an excess of

$TamB_{DUF490}$ (20-100 μM) in a 1:1 ratio produced a rapid fluorescence quenching phase which slowly declined over a 5 s period.

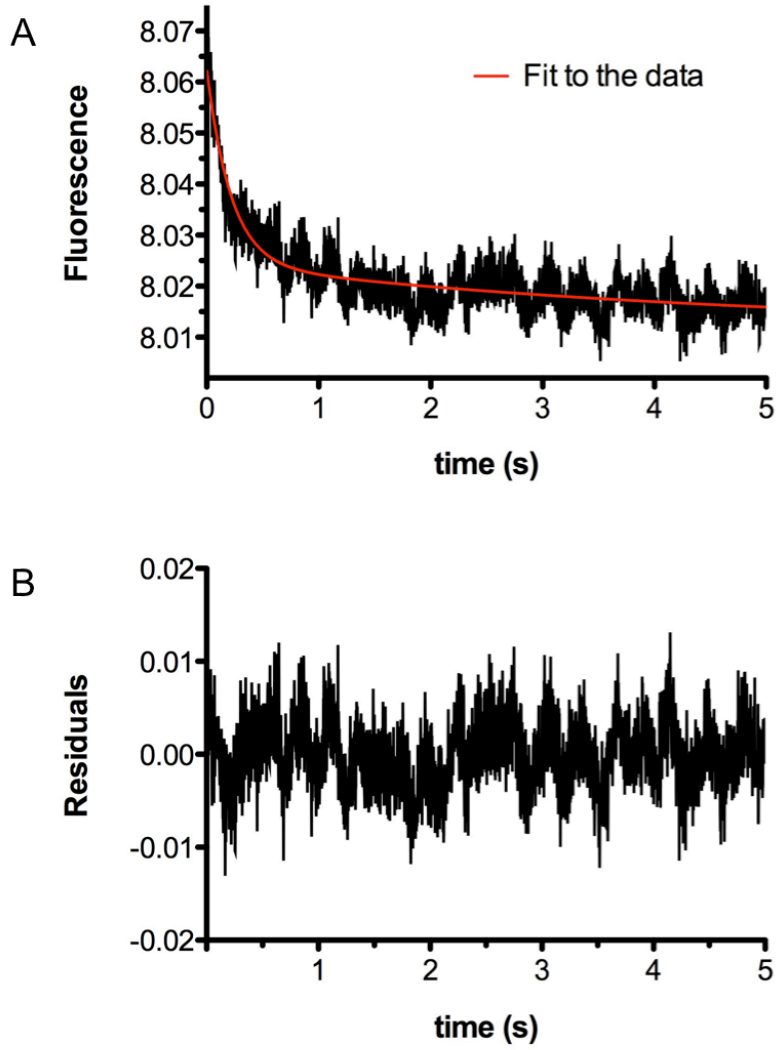


Figure 5-10 Association curve for $TamA_{POTRA}$ and $TamB_{DUF490}$ (A) and residuals (B). The concentration of $TamA_{POTRA}$ was 10 μM and concentrations of $TamB_{DUF490}$ ranged from 20 – 100 μM prior to injection. Data were fitted with a double-exponential model yielding the rate constants k_{obs1} and k_{obs2}. All experiments were performed in 0.2 M sodium phosphate, pH 7.5, 0.015% Triton X-100.

No change in fluorescence was observed during the mixing of the individual proteins alone ($TamA_{POTRA}$ at 5 μM and $TamB_{DUF490}$ at 200 μM). The curve was fitted with a double-exponential rate equation, since a single exponential function produced a poor fit, and

the k_{obs1} and k_{obs2} rate constants were derived (Figure 5-10). No deviating trends were observed in the residual plot. Plotting k_{obs1} versus the experimental $TamB_{DUF490}$ concentration range produced a linear plot with some minor concentration-dependent variation observed in the value of k_{obs2} (Figure 5-11). The linearity of the k_{obs1} variation with increasing $TamB_{DUF490}$ concentration indicated that it represents the rate constant for the bimolecular association step. The slope of the line represents the association rate constant (k_{on}) and the y-intercept the dissociation rate constant (k_{off}) (Figure 5-11 A). The bimolecular association rate for the binding reaction was calculated to be $1.1 \pm 0.25 \times 10^5$ M^{-1} s^{-1} (R^2=0.87) with the off-rate constant being 0.492 s^{-1}. The equilibrium dissociation constant (K_d) derived from these parameters (k_{off}/k_{on}) was calculated to be 4.45 μM, which is in accord with the calorimetric measurements of the $TamA_{POTRA}$-$TamB_{DUF490}$ association (Figure 5-3). The limited variation observed in k_{obs2} with increasing $TamB_{DUF490}$ concentration may indicate the presence of conformational heterogeneity in the binding sites of one of the associating partners. However, the protein concentration used in this experiment may not be high enough to monitor the effect on k_{obs2} during complex formation.

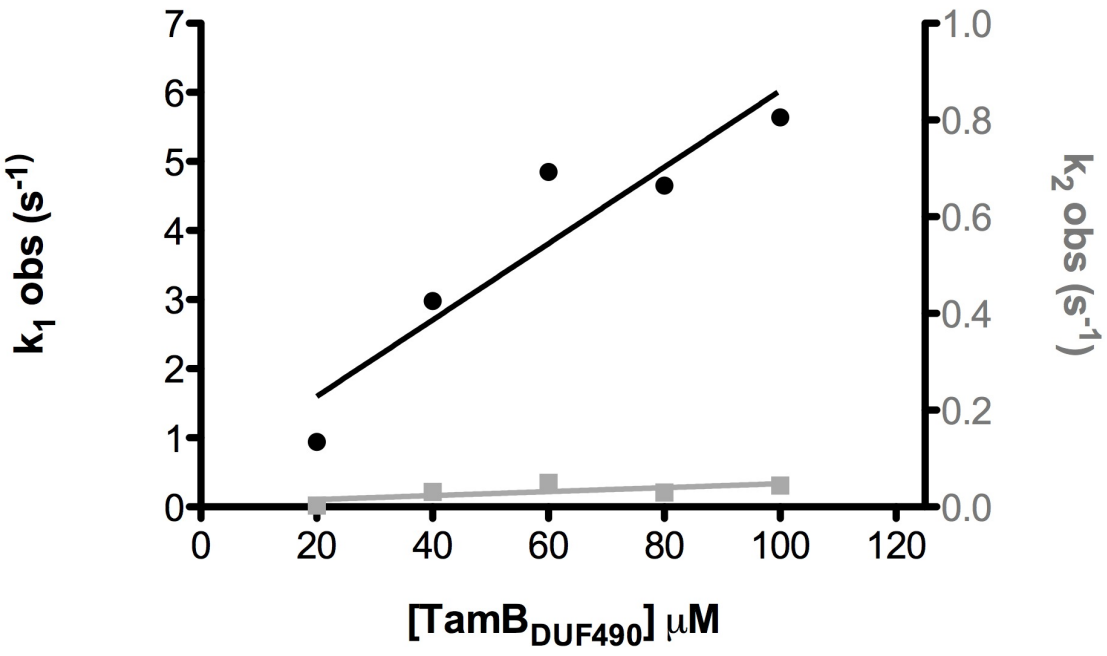


Figure 5-11 Plot of k_{obs1} and k_{obs2} versus increasing $TamB_{DUF490}$ concentrations. k_{obs1} was taken as the bimolecular association constant. k_{obs2} showed only minor concentration-dependent variation in the reactions which may be attributed to the encounter complex/conformational changes in $TamA_{POTRA}$.

5.2.5 *TamB binding to $TamA_{POTRA}$ leads to reduced tumbling of the POTRA1 domain as probed by NMR spectroscopy*

Kinetic experiments between $TamA_{POTRA}$ and $TamB_{DUF490}$ indicated that TAM complex formation could involve conformational re-arrangements upon molecular association. Moreover, solution characterisation of TamA POTRA domains revealed some degree of inter-domain mobility within the domains. Therefore, to assess whether the conformational changes observed using stopped-flow could be, at least partially, attributed to the POTRA domains, TamA-TamB binding was explored using NMR spectroscopy. Addition of excess $TamB_{DUF490}$ (5-fold excess) to ^{15}N-labelled $TamA_{POTRA1}$ does not result in any chemical shift changes in the domain, indicating that no interactions take place between these two molecules, which is in agreement with the ITC data presented in the previous chapter (Figure 5-5). However, titration of excess of $TamB_{DUF490}$ into ^{15}N-labelled $TamA_{POTRA12}$ causes significant changes to the chemical shifts in the HSQC spectrum. Notably, the cross-peaks originating from the POTRA1 domain

become extensively broadened in the presence of excess $TamB_{DUF490}$. This is consistent with the formation of a large macromolecular complex consisting of $TamA_{POTRA12}$ (19 kDa) and $TamB_{DUF490}$ (42 kDa) associated with Triton X-100 micelles (~ 95 kDa) (Figure 5-12). Titration of Triton X-100 alone into $TamA_{POTRA12}$ up to a concentration of ten times the critical micelle concentration does not lead to significant changes to the $TamA_{POTRA12}$ spectra (Figure 5-14). Therefore, this spectral broadening of POTRA1 chemical shifts can be ascribed to conformational changes in response to $TamB_{DUF490}$ association.

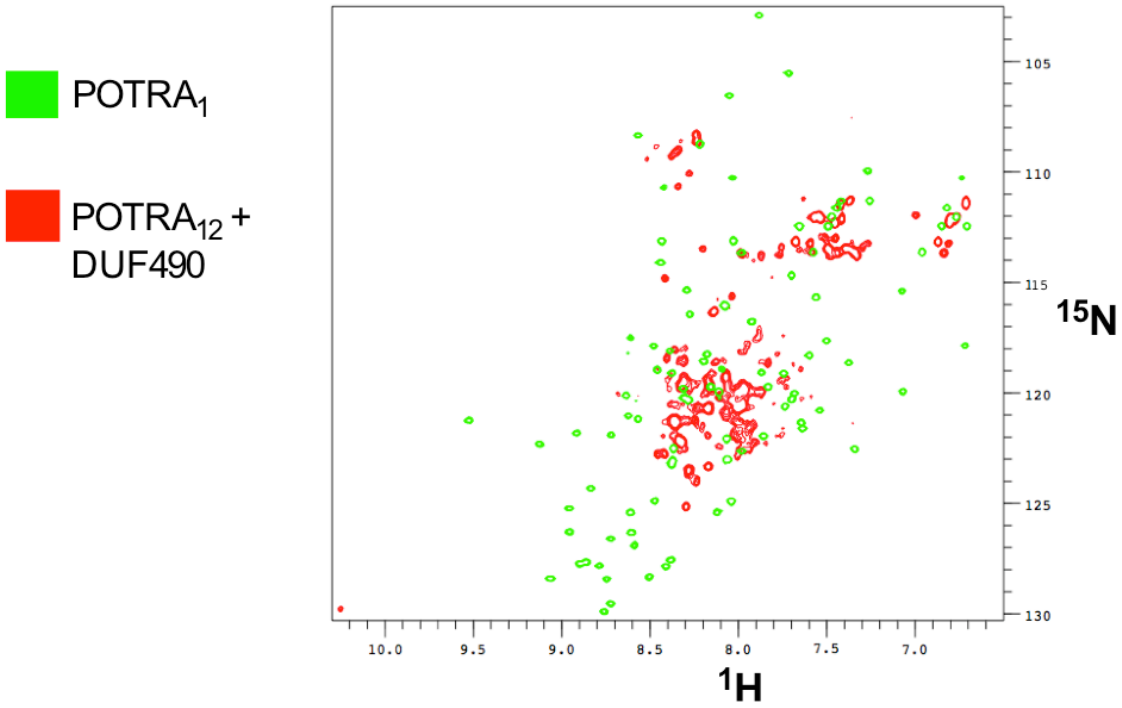


Figure 5-12 $TamB_{DUF490}$ affects the solution dynamics of POTRA1 upon associating with $TamA_{POTRA12}$. Overlay of the HSQC spectra of ^{15}N-labelled $TamA_{POTRA1}$ with a complex of $TamB_{DUF490}$ and ^{15}N-labelled $TamA_{POTRA12}$ shows effects of $TamB_{DUF490}$ on the dynamics of POTRA1. Experiments were performed in 0.1 M sodium phosphate pH 6.9, 298 K.

In addition, chemical shift changes are observed in response to $TamB_{DUF490}$ within peaks originating from POTRA2 as illustrated by the chemical shift in the indole region of the ^{1}H-^{15}N HSQC spectra that originates from the side chain of the sole $TamA_{POTRA12}$ tryptophan, Trp179, located in POTRA2 (Figure 5-13).

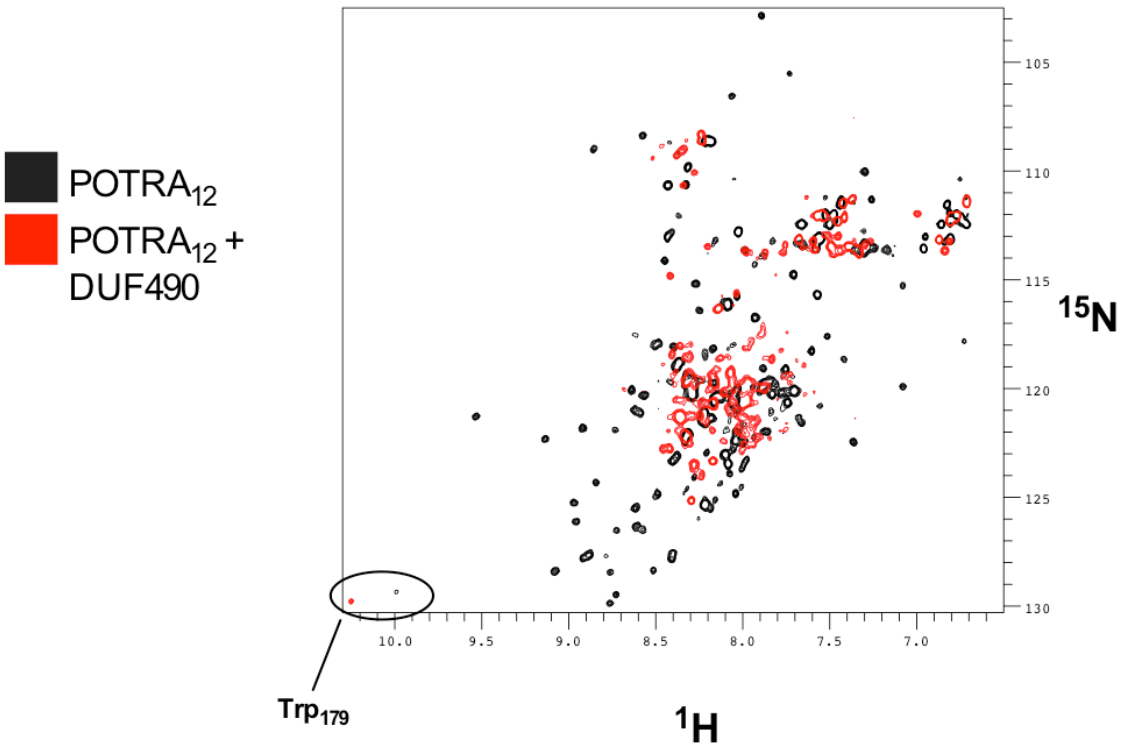


Figure 5-13 $TamA_{POTRA12}$-$TamB_{DUF490}$ interactions probed by NMR spectroscopy. Overlay of ^{1}H-^{15}N HSQC spectrum of $TamA_{POTRA12}$ and $TamA_{POTRA12}$ + $TamB_{DUF490}$ emphasises the effect on dynamics by $TamB_{DUF490}$. Highlighted is the indole resonance belonging to Trp179 located in POTRA2 and the chemical shift arising from its perturbation by $TamB_{DUF490}$. Experiments were performed in 0.1 M sodium phosphate pH 6.9, 298 K

Moreover, the number of cross-peaks and their spread is increased in the presence of $TamB_{DUF490,}$ which suggests that some degree of folding takes place within the POTRA2 domain as a result of the interaction with $TamB_{DUF490}$. A control experiment was performed to ensure that the linewidth broadening of POTRA1 was a result of DUF490 interactions and not interference from the Triton X-100 micelles. 0.15-0.2% Triton X-100 (10x the c.m.c. value for this detergent) was mixed with ^{15}N-labelled $TamA_{POTRA12}$ and HSQC spectra were acquired. No large-scale changes in the cross-peaks of $TamA_{POTRA12}$ were detected (Figure 5-14). Minor perturbation was seen in the chemical shifts corresponding to residues from both POTRA1 and partially folded POTRA2. Moreover, no broadening of the chemical shifts from domain 1 was evident.

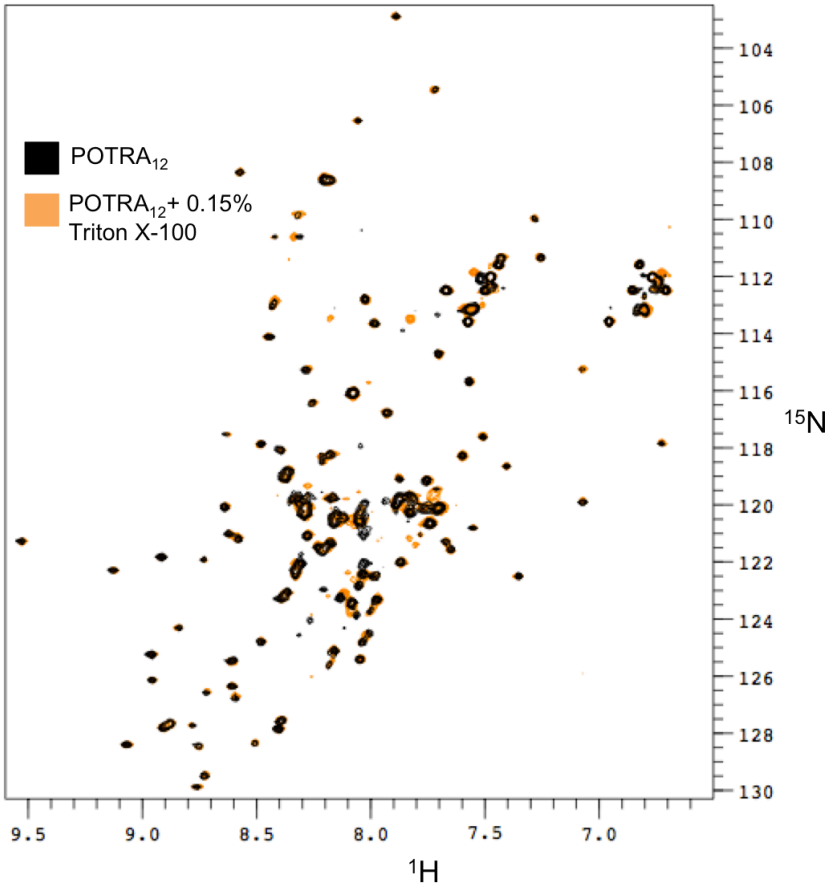


Figure 5-14 Addition of Triton X-100 (10x the c.m.c.) does not lead to large spectral pertubation of $TamA_{POTRA12}$. 50 μM of ^{15}N-labelled $TamA_{POTRA12}$ in 0.1 M sodium phosphate, pH 6.9 was mixed with excess Triton-X100 (0.2% final concentration).

Next, to ensure that the spectral perturbation of POTRA1 as a result of $TamB_{DUF490}$ binding to $TamA_{POTRA12}$ was not related to the molten globule-like property of POTRA2, full-length ^{15}N-labelled $TamA_{POTRA}$ was utilised in the binding assay. TROSY spectra were acquired, but due to sample limitation and instability, the quality of the collected spectra was very poor. Not all cross-peaks were resolved due to under-sampling of the data (Figure 5-15 blue). Nevertheless, upon addition of $TamB_{DUF490}$ the cross-peaks corresponding to the first POTRA domain were again broadened, suggesting that even in the context of the full-length $TamA_{POTRA}$ interaction with $TamB_{DUF490}$ leads to the structural perturbation of POTRA1. Moreover, since there are two Trp residues in the full-length $TamA_{POTRA}$ construct (POTRA2 and POTRA3), both are perturbed during the binding reaction. This finding is in accord with the reduction in the K_d measured by the ITC experiments (Figure

5-4). Therefore, some contribution from the third POTRA domain to the association with $TamB_{DUF490}$ is expected.

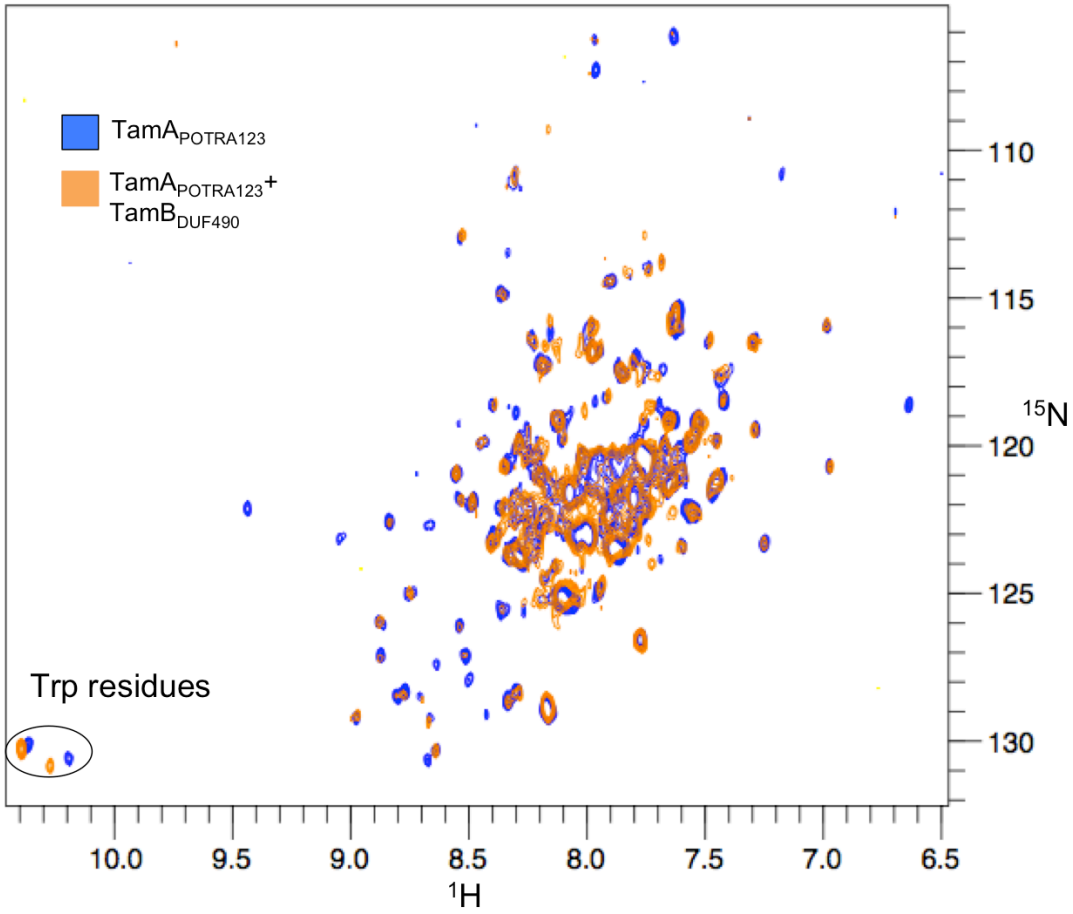


Figure 5-15 $TamB_{DUF490}$ leads to structural perturbation of POTRA1 in the full-length $TamA_{POTRA}$. TROSY spectra were acquired for a 60 μM sample of ^{15}N-labelled $TamA_{POTRA}$. Approximately 180 μM $TamB_{DUF490}$ was used for the binding reaction. Experiments were performed in 0.1 M sodium phosphate, 0.1 M NaCl pH 6.9, 293 K.

5.3 Discussion

TAM complex has previously been shown to span the periplasmic space through the association of an IM protein TamB and an OMP TamA (138). Further insights into this interaction have revealed that the conserved C-terminal domain of unknown function, DUF490, was responsible for mediating this interaction. This chapter was mainly

concerned with the biophysical and biochemical characterisation of interactions between the POTRA domains of TamA and the DUF490 domain from TamB *in vitro.*
There are currently no data to describe the structural organisation of TamB, except for the C-terminal presence of the DUF490 domain, yet the cellular localisation of this domain (i.e. periplasmic versus cytoplasmic) is currently unknown. By relying on the bioinformatic prediction of the C-terminal transmembrane helix, it was concluded that at least the last 80 amino acids of TamB must be periplasmic. Additionally, previously published results suggest that $TamB_{DUF490}$ lacking the last 80 amino acids does not associate with TamA (138). In this chapter, direct binding interactions between $TamB_{DUF490}$ and $TamA_{POTRA}$, as well as the $TamB_{C80}$ fragment were measured. These experiments have shown that the 80 C-terminal amino acids of $TamB_{DUF490}$ are necessary for binding between TamA and TamB. Differences in the thermodynamic characteristics of association of full-length $TamB_{DUF490}$ and $TamB_{C80}$ for $TamA_{POTRA}$ are suggestive of structural re-arrangements within $TamB_{DUF490}$ upon TAM complex formation. In order to try to localise the POTRA domain in TamA responsible for complex formation, POTRA3 and POTRA23 were deleted from $TamA_{POTRA}$. The $TamA_{POTRA1}$ domain produced no heats of interaction suggesting that POTRA1 domain in isolation is not crucial for TAM complex formation, with $TamA_{POTRA2}$ being the major binding determinant of the three POTRA domains of TamA and POTRA3 having a minor contribution to the overall free energy of binding.

ITC and stopped-flow experiments revealed that the affinity of the interaction between TamA and TamB is in the low micromolar range. This type of complex can therefore be described as transient and reversible (K_d range in milimolar to micromolar), in contrast to stronger, more stable protein complexes with extensive life-times (K_d ranges in nanomolar to femtomolar). The reversibility of TAM complex formation could be physiologically relevant to the functioning of the complex and its individual components. Comparison of the computed isoelectric points of the two domains (pI of $TamB_{DUF490}$ = 4.52, $TamA_{POTRA}$ = 8.87) is suggestive of a charge complementarity between the two binding regions, however inspection of the thermodynamic properties of complex formation indicate that entropy could be the driving force behind the global binding reaction. In-depth studies of encounter complexes have suggested that long-range

electrostatic forces are key for encounter complex pre-orientation without major desolvation of the binding surfaces (194). Therefore, associations between non-polar segments from both molecules might be important for the transition from the encounter complex to the final step of TAM complex formation. Favourable desolvation of hydrophobic surfaces, as could be case in TamB, can help outweigh any conformational restrictions within the interacting polypeptides that would otherwise disfavour the entropy of the reaction.

Even more interestingly, the conformation of TamA POTRA domains, or at least POTRA1, seems to be modulated by the partner protein TamB, as shown by the NMR binding studies. The conformational change in POTRA1 from a free tumbling state to that of a rigid complex could have functional implications in the role of transport of AT passenger domains across the OM by TamA. Since the main TamB binding site lies within the second POTRA domain, the effect on POTRA1 dynamics could be allosteric. No such structural modulation has been observed for the BamA POTRA domains, however it is possible that a similar phenomenon exists. For example, the main hinge point in BamA POTRAs is found between POTRA2 and POTRA3; POTRA3 has also been shown to be the main binding site of BamB. Although the precise function of BamB is currently not known, functional studies indicate that it is involved in OMP folding (55). It is therefore plausible that BamB could be involved in the modulation of BamA POTRA conformations, however this observation needs further corroboration.

The crystal structure of TamA POTRA together with the C-terminal peptide shows that the hydrophobic peptide makes several contacts with the linker between POTRA1 and POTRA2, and although speculative, it could reflect the locking of POTRA1 in the presence of DUF490. This would imply that the currently available crystal structure could represent the 'TamB-bound' conformation of the domains, which would explain the discrepancy between the SAXS and crystallographic data. It is currently not known whether TamA and TamB function together as a complex in AT biogenesis or whether the individual molecules contribute to the process in a sequential mechanism of action. It is possible to envisage a mechanism whereby the TamB-bound TamA POTRA domains possess the correct conformation needed for substrate transport and association.

POTRA domains from BamA have been hypothesised to utilise β-augmentation as a means of forming contacts with substrates during β-barrel OM insertion (30,96). Primary evidence for this mechanism stems from the crystal structue of POTRA1-4 from BamA, where several domains were seen to 'lend' a β-strand to a neighbouring molecule within the crystal-packing lattice. Additionally, Knowles *et al.* utilised a chemical shift assay using POTRA1-2 and β-sheet peptides from PhoE, a substrate for the BAM complex, to confirm that transient interactions between POTRA and putative β-strands takes place (58). However, no direct evidence for this mechanism in BamA has been presented yet, inasmuch as it is not known whether the augmenting β-strands require pre-existing secondary structure, induced by the neighbouring polypeptide chain, or whether the POTRA domains induce β-strand formation *de novo* in a nucleating fashion. To date, no structural information is available to describe whether interactions between BamA POTRAs and other members of the BAM complex occur by β-augmentation. The high-resolution crystal structure of BamB shows that the protein has a β-propeller fold (66) and could utilise β-augmentation to form the BamA-BamB complex. On the contrary, BamD is an all-α-helical protein. Therefore, its mechanism for associating with POTRA5 of BamA must be fundamentally different.

5.4 Conclusion

In this chapter interactions between TamA and TamB have been measured using several biophysical approaches. The main binding sites have been localised to the extreme C-terminus of TamB and the second POTRA domain of TamA. Association between the two proteins appears to be entropically driven with a minor favourable enthalpic element and proceeds via a transient intermediate. NMR spectroscopic studies of the binding reaction suggest that DUF490 is involved in the structural modulation of POTRA domains. These biophysical studies pave the way for potential targeting of the TAM complex in terms of therapeutic approaches for anti-virulence compounds and peptides.

6 Recognition of the Ag43 passenger domain by TamA and TamB

All NMR experiments and sequence-specific backbone assignments were carried out under the supervision of Dr Brian O. Smith.

CD experiments were carried out in collaboration with Dr. Sharon Kelly

6.1 Introduction

All ATs are synthesised in the cytoplasm as pre-proproteins consisting of a signal peptide (SP), a passenger domain and a β-domain and utilise the Sec translocation system in order to reach the periplasmic space (101). In the periplasm, the SP is cleaved by the leader peptidase to produce the pro-protein. Once the pro-protein is secreted outside the cell it undergoes (auto)catalytic cleavage of the passenger domain (119). The α-domain then remains either non-covalently attached to the β-domain or is secreted into the extracellular space. The mechanism behind α-domain secretion is only partially understood. It has been established that the TAM complex contributes to the transport of AT passenger domains across the OM, however the precise mechanism of its function has yet to be elucidated (138). Previously identified as a TAM substrate, Ag43 is an adhesin AT that is involved in bacterial cell-cell adhesion, bacterial autoaggregation and biofilm formation and was shown to accumulate as a full-length pro-protein in the periplasmic space in a Δ*tamAB* mutant of *E. coli* K-12 (138,195). Additionally, cell surface delivery of the α subunit of the Ag43 passenger domain was greatly diminished in Δ*tamAB* cells (138). The recently solved crystal structure of the Ag43 passenger domain revealed an L-shaped β-helix with a propensity to self-associate into homodimers, as shown using complementary biophysical methods (Figure 6-1) (130). In addition to the TAM complex, AT biogenesis has been shown to involve the common periplasmic chaperones SurA, Skp and DegP as well as the general OMP biogenesis machinery, the BAM complex.

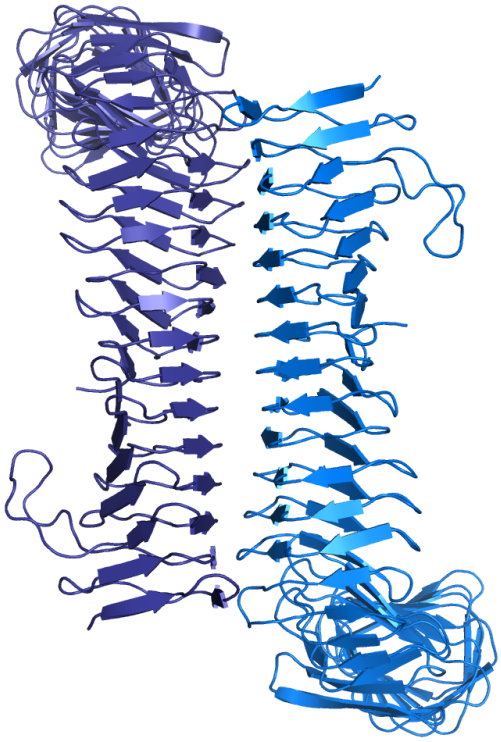


Figure 6-1 High-resolution crystal structure of Ag43 passenger domain suggests mechanisms for the auto-aggregation function of this adhesin. Recombinantly expressed Ag43 passenger domain exhibited concentration-dependent dimerisation in solution, which has been shown to drive bacterial cell autoaggregation *in vivo*. The dimer interface is extensive and consists of loop-loop interactions along the entire length of the passenger domain (130) **(**PDB ID: 4KH3).

Binding studies by Ruiz-Perez *et al.* have shown that the periplasmic chaperones SurA and DegP are able to interact with the passenger domain of EspP (115). Additionally, Ieva *et al*. have shown that a stalled passenger domain intermediate of EspP associates transiently with several other periplasmic chaperones such as Skp, BamB and BamA through its passenger domain as well as its β-domain (Figure 6-2) (60). SurA has also been shown to associate with the BamA POTRA domains, thus providing a link between the translocating periplasmic intermediates and transport via the BAM complex (59).

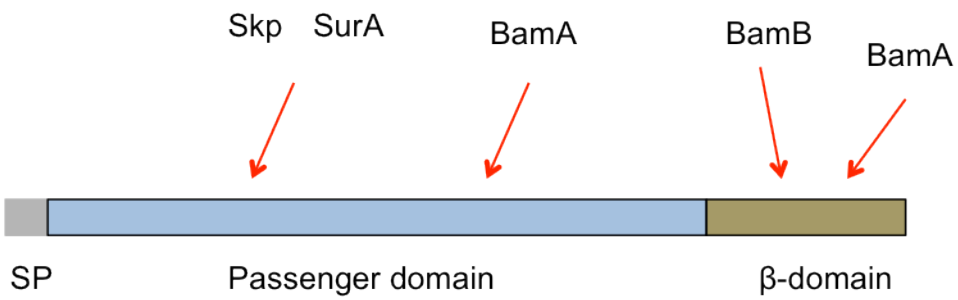


Figure 6-2 Experimentally detected interactions between the periplasmic chaperones and EspP. SurA and Skp have been shown to associate in the proximity of each other with the EspP passenger domain, with Skp association taking place in the early stages of the periplasmic life-cycle of the passenger domain, with SurA contributing to the later stages of passenger domain translocation (114). BamA associates with both the passenger domain and the β-domain, suggesting that it might aid in the translocation of the passenger domain across the OM (60,114).

The formation of a cell envelope-spanning complex between TamA and TamB and its contribution to AT biogenesis suggests that both proteins may act as additional key players involved in AT biogenesis and translocation. The role of TamA, being a member of the Omp85 family of transporter proteins, could reside with the translocation of the passenger domain across the OM, whilst TamB could act as a chaperone in the delivery of AT passenger domains to the TamA barrel, playing a role analogous to that played by the soluble chaperones Skp and SurA in the delivery of integral membrane β-barrel domains to the BAM complex (60).

Protein translocation across the IM usually occurs in an N- to C- terminal fashion. Therefore, the N-terminus of ATs is the first polypeptide segment to be found in the periplasmic space during AT biogenesis. Several studies have shown that OM translocation of AT passenger domains occurs directionally in a C- to N- terminal manner and is an ATP-independent process that is decoupled from the proton-motive force (131,133). Folding and secretion studies of the AT pertactin have shown that the N-terminal region of the passenger domain is structurally unstable, in contrast to the C-terminal region which is highly stable and is resistant to protease degradation *in vitro.* It was therefore suggested that the C-terminal region of the passenger domain could act as a folding scaffold for the rest of the domain during its translocation across the OM. Moreover, equilibrium unfolding kinetics indicate that the overall folding of the β-helical

passenger domain is slow, a property that might aid in maintaining its translocation-competent conformation in the periplasm (127). This chapter examines interactions of TamA and TamB with the N-terminal region of the Ag43 passenger domain.

6.2 Results

6.2.1 *Overexpression and purification of Ag43$_{N168}$*

Passenger domain secretion of several ATs is facilitated by the periplasm-spanning TAM complex, consisting of TamA and TamB. In order to examine the nature of interactions of TamA and TamB with substrate passenger domains, recombinant production of the passenger domain of the substrate AT Ag43 was undertaken. Over-expression and purification of several constructs was attempted; however, only one construct, representing the N-terminal 168 residues (hereby referred to as Ag43$_{N168}$), proved to be soluble and could be readily purified in sufficient yield for detailed biophysical analyses.

The coding sequence for the N-terminal region of Ag43, residues 56-219 (calculated M_w 18.1 kDa), was cloned into pET21a to encode a C-terminally His$_6$-tagged protein. For recombinant protein production this plasmid (pAg43$_{N168}$) was transformed into BL21(DE3) and 5 litres of culture were grown at 37°C until OD_{600} 0.4-0.6 was reached and 1 mM IPTG was added. Cells were then left to grow at 30°C for 6-7 hours and harvested by centrifugation. The recombinant protein was purified using Ni^{2+}-affinity chromatography, followed by gel-filtration chromatography on a Superdex S75 column. The purity of Ag43$_{N168}$ was assessed using 15% SDS-PAGE and was seen to be pure (>95%)(Figure 6-3 A,B). The protein eluted as a single peak by gel-filtration and SDS-PAGE revealed a single band migrating as an appoximately 24 kDa protein. Abnormal electrophoretic mobility of Ag43$_{N168}$ could be ascribed to its acidic nature (pI = 5.14) and incomplete coating with SDS resulting from repulsions between the detergent and negatively charged protein.

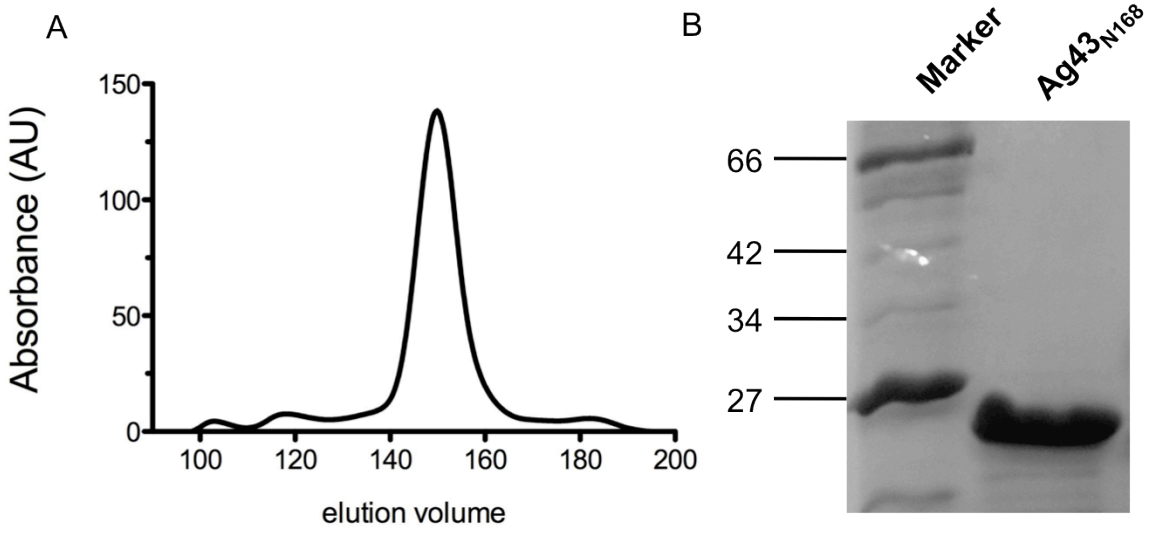


Figure 6-3 Purification of $Ag43_{N168}$. (A) Size-exclusion profile of $Ag43_{N168}$ shows that the protein behaves as a single species in solution with an elution volume of 155 ml. (B) SDS-PAGE showing the purity of $Ag43_{N168}$, and an abnormal electrophoretic mobility on a 15% gel.

In order to confirm the identity of the purified $Ag43_{N168}$ the protein from an SDS-PAGE band was analysed by N-terminal sequencing. This analysis gave the expected sequence corresponding to the first 5 amino acids (DIVVH) of $Ag43_{N168}$. Purified $Ag43_{N168}$ was analysed by electron-spray ionisation mass spectrometry (ESI-MS) in order to determine whether the experimentally determined M_w of the purified polypeptide corresponded to its amino acid-derived M_w. The results revealed a single, dominant peak with a measured mass of 18133 Da, which is close to the M_w calculated from the amino acid protein sequence (18132 Da) (Figure 6-4). Once the identity and composition of $Ag43_{N168}$ was confirmed, the protein was characterised structurally and biophysically.

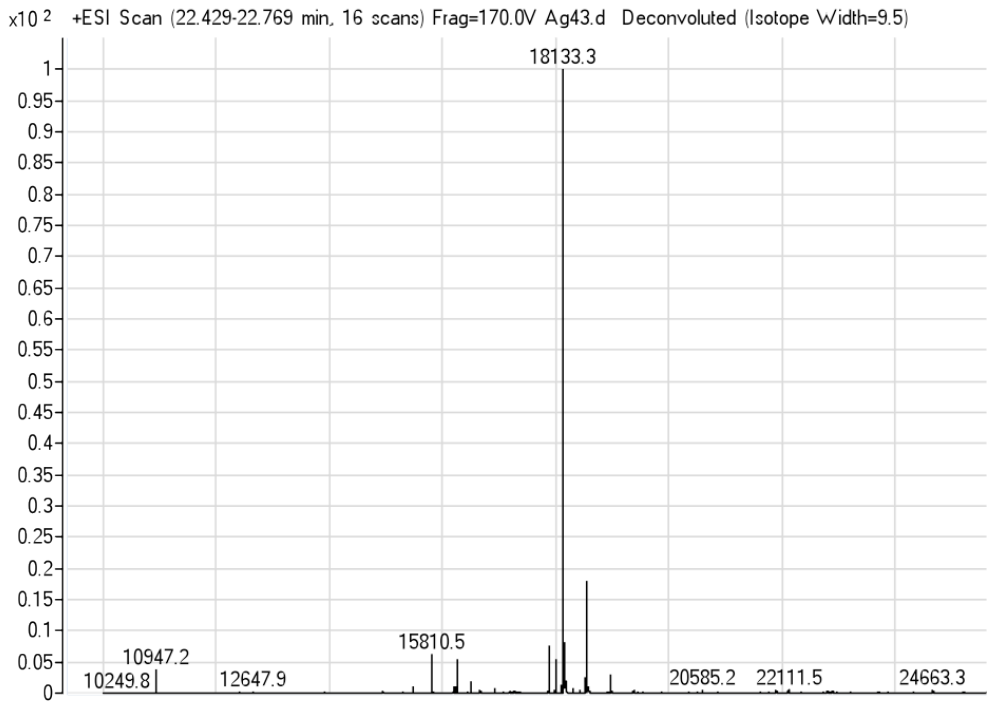


Figure 6-4 ESI-MS analysis of purified $Ag43_{N168}$ confirms the expected M_w. A single dominant peak with an observed M_w of 18133 Da is seen, close to the expected molecular mass of 18132 Da.

6.2.2 *Structural and biophysical characterisation of $Ag43_{N168}$*

The recently solved crystal structure of full-length Ag43 passenger domain (PDB ID: 4KH3) from an *E. coli* UPEC strain UTI89 indicates that the protein can form homodimers (130). One of the observed dimers was formed through inter-molecular contacts of the N-terminal regions of two symmetry-related molecules (Figure 6-5) and may represent a potential binding interface, however this remains to be proven. In order to characterise the oligomeric state of the $Ag43_{N168}$ construct, sedimentation velocity and sedimentation equilibrium studies were conducted at a range of $Ag43_{N168}$ concentrations.

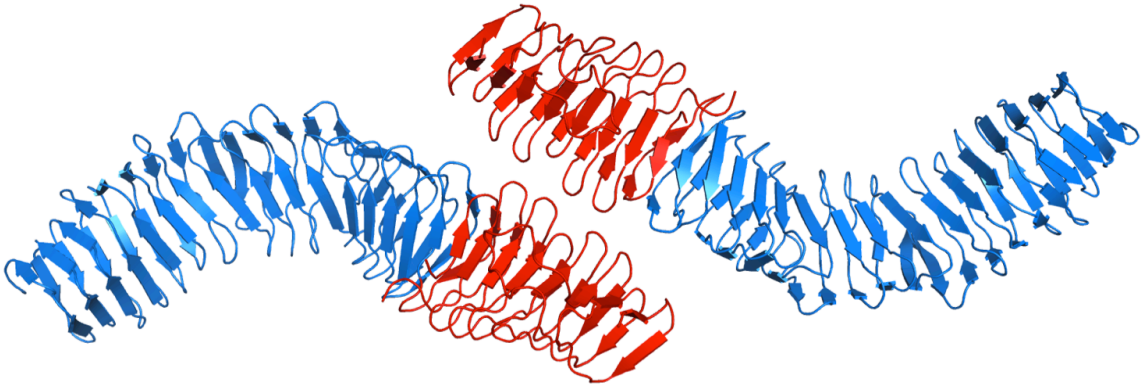


Figure 6-5 One of the putative dimerisation interfaces (residues 54-220) between two symmetry related molecules. Highlighted in red is the polypeptide region corresponding to the recombinantly expressed $Ag43_{N168}$ construct. Since the dimerisation observed *in crystallo* could be reflective of the multimeric nature of the protein observed *in vivo,* the purified $Ag43_{N168}$ could exist as a dimer (PDB ID: 4KH3).

In order to investigate the oligomeric state of $Ag43_{N168}$, analytical ultracentrifugation experiments were carried out. Sedimentation velocity (SV) reports on the sedimentation coefficient of the protein, a value that is related to the particle's molecular mass and shape as it moves through solution under a high centrifugal force. The sedimentation coefficient (s) is directly proportional to the molecular mass and inversely proportional to the frictional coefficient, which is related to the molecular shape in solution. Moreover, by inspecting the continuous sedimentation coefficient distribution it is possible to determine the number of distinct species (monomers, dimers etc.) present in a sample. SV experiments carried out for a range of $Ag43_{N168}$ concentrations indicated that the molecule is monodisperse without any evidence of the formation of higher molecular weight species (Figure 6-6A). The determined sedimentation coefficient for the protein was 1.25 S, whereas most proteins of a similar M_w which are globular in shape (e.g. lysozyme and myoglobin) typically have s values closer to 2 S (196).

Sedimentation equilibrium (SE) studies are aimed at determining the protein M_w based on the interplay between particle's sedimentation and diffusion properties, without any experimental contribution from the particle's shape. Once the equilibrium between the sedimentation and diffusion forces is established a concentration boundary is created, which allows the estimation of the M_w of a protein from first principles. SE of $Ag43_{N168}$ showed that the molecule is monomeric with a M_w of 17 kDa, close to the expected M_w of

18.1 kDa for monomeric $Ag43_{N168}$. Having determined the solution M_w of $Ag43_{N168}$ and confirmed its monomeric nature, a more detailed analysis of the hydrodynamic properties of $Ag43_{N168}$ was performed.

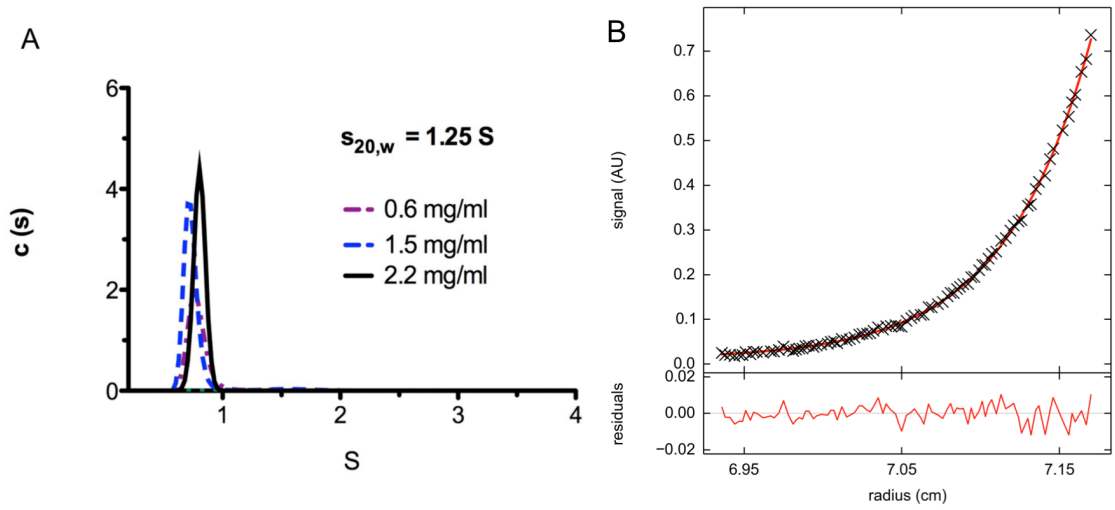


Figure 6-6 Analysis of $Ag43_{N168}$ by analytical ultracentrifugation. (A) Continuous c(s) distribution profile of $Ag43_{N168}$ indicates that the expressed construct is monodisperse and monomeric in solution. Additionally, it possesses a very low sedimentation coefficient ($s_{20,w}$ = 1.25 S) which, for a protein with a molecular weight of 18.1 kDa, suggests a high frictional ratio. (B) Sedimentation equilibrium data of $Ag43_{N168}$ (black crosses) reveal that the molecule exists as a monomer with a measured M_w of 17 kDa. SE data were fitted with a single species model (red line). Experiments were performed in 0.1 M sodium phosphate buffer, pH 7.5.

The hydrodynamic properties of $Ag43_{N168}$ were predicted *in silico* using the program US-SOMO (156), which calculates hydrodynamic parameters of high-resolution 3D models of proteins (here the crystal structure of Ag43 was used, taking the first 168 amino acids of the passenger domain model, PDB ID: 4KH3). The computed sedimentation coefficient of the $Ag43_{N168}$ model based on the crystal structure yielded a value of 1.9 S, which deviates significantly from the experimentally determined $s_{20,w}$ value. Possible reasons for this difference could lie in the structural discrepancy between the high-resolution model of Ag43 and the purified $Ag43_{N168}$, arising from structural flexibility and high particle anisotropy, which would lead to an increased frictional coefficient. Overall, the AUC data indicate that structurally, the purified construct of $Ag43_{N168}$ is not representative of the

corresponding region in the 3D model of the passenger domain seen in the crystal structure.

$Ag43_{N168}$ was analysed using far-UV circular dichroism (CD) to quantify the secondary structure composition of the molecule, which was expected to be β-strand rich. Estimates from the far-UV CD analysis indicates that $Ag43_{N168}$ consists of 61% random-coil and 29% β-strand with minor contribution from α-helical segments (~10%) (Figure 6-7).

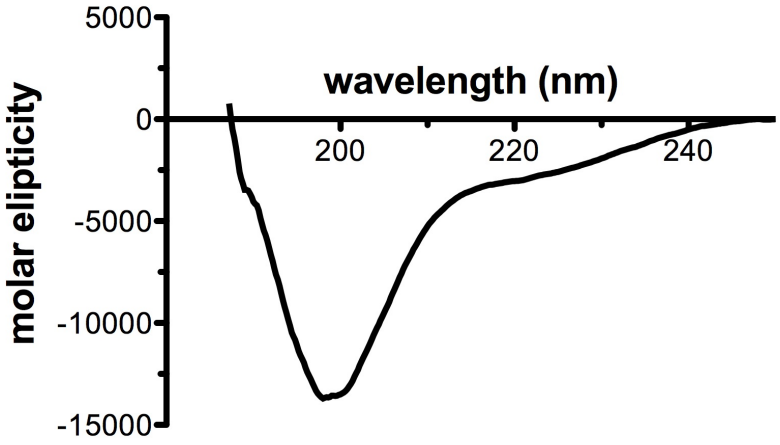


Figure 6-7 Far-UV CD spectrum of $Ag43_{N168}$ indicates that 60% of the polypeptide has no regular secondary structure elements, with some minor β-strand character. Spectra were acquired for 39 μM $Ag43_{N168}$ in 0.1 M sodium phosphate buffer, pH 7.5.

This suggests that the purified $Ag43_{N168}$ construct deviates from the high-resolution model of the Ag43 N-terminus in terms of its secondary structure content.

6.2.3 *$Ag43_{N168}$ lacks a well-defined tertiary structure in solution as probed by NMR spectroscopy.*

The high proportion of non-regular secondary structure in $Ag43_{N168}$, along with its high frictional coefficient in solution, as determined by SV experiments, indicates that the polypeptide could exhibit a high degree of intrinsic disorder or the presence of numerous highly flexible regions. In order to investigate the solution structure of $Ag43_{N168}$ further,

^{15}N-labelled Ag43$_{N168}$ was produced for NMR studies. NMR provides solution structural information about proteins at the atomic level. Basic spectra such as HSQC spectra provide a wealth of information about the structure of the molecule such as the level of foldedness, which is visualised by ^{1}H chemical shift dispersion, and the presence of flexible regions in the polypeptide which are revealed by changes in the peak intensity with varying temperature. ^{1}H-^{15}N HSQC spectra were acquired using 50-100 μM Ag43$_{N168}$ at a range of temperatures (4-25 °C). At 25°C the HSQC spectrum of Ag43$_{N168}$ shows broad peaks of heterogenous linewidth with poor dispersion in the ^{1}H dimension (Figure 6-8 A). The majority of the visible cross-peaks exhibit significant overlap (Figure 6-8A). Variation in peak intensities is attributed to significant chemical exchange with the bulk solvent water, possibly due to large solvent exposure of the polypeptide backbone (197). These spectral characteristics are highly suggestive of a molecule that lacks a well-defined tertiary structure and numerous flexible regions in its polypeptide chain. Reducing the temperature of the sample to 4°C resulted in improvement of the HSQC spectra by reducing peak overlap with the majority of peaks being of homogenous linewidth. At 4°C most peaks can be resolved due to reduced exchange rates of amides with the solvent water molecules, with the number of backbone amide peaks (151 peaks), excluding the NH_2 and indole regions of the spectrum, in the HSQC corresponding closely to the number of amino acids within the construct (176 residues).

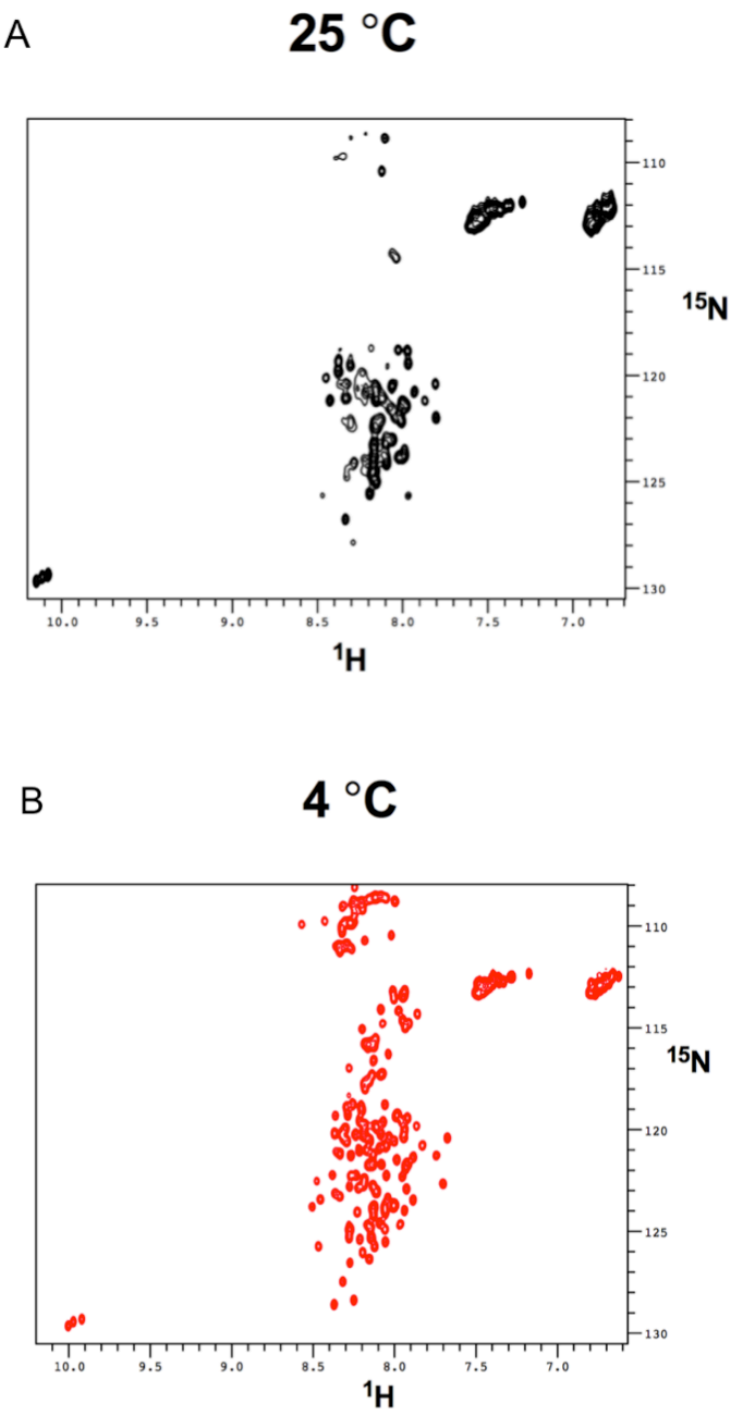


Figure 6-8 The solution structure of $Ag43_{N168}$ probed by NMR spectroscopy suggests that the molecule exists as an intrinsically disordered polypeptide. (A) ^{1}H-^{15}N HSQC spectra of $Ag43_{N168}$ recorded at 25°C (A) and 4°C (B). The narrow spectral range of cross-peaks is representative of an intrinsically disordered protein. This idea is further strengthened by the improvement of spectral quality with a reduction in temperature, which suppresses the molecular tumbling into an exchange regime that is observable by NMR. Spectra were acquired for 200 μM ^{15}N-labelled $Ag43_{N168}$ in 50 mM sodium phosphate buffer, pH 6.9.

The combined AUC, CD and NMR data confirm that $Ag43_{N168}$ exists predominantly as polypeptide in a state of intrinsic disorder, rather than the expected β-helical conformation. Intrinsically disordered proteins (IDPs) are conformationally heterogeneous molecules, which resemble unfolded polypeptide chains of globular proteins. However, not all IDPs exist as extended polypeptide chains and some local structural elements, even though temporal, infer some polypeptide chain compaction (198). IDPs sample a continuum of conformations due to a lack of dihedral restraints of secondary structure elements about amino acid positions and can interchange between compact and extended conformations, usually represented by ensembles on millisecond timescales (199,200). Nevertheless, the far-UV CD measurements indicate that some secondary structure elements are present in $Ag43_{N168}$.

6.2.4 *SAXS analysis of $Ag43_{N168}$ reveals an extended, disordered particle in solution*

In order to gain further insight into the solution structure of $Ag43_{N168}$, SAXS data were collected for a range of concentrations (0.7-4.5 mg ml^{-1}). Consistent with the NMR data, the shape of the raw scattering curve for $Ag43_{N168}$ is featureless, which is very typical of IDP-like proteins (Figure 6-9A). The experimentally derived radius of gyration, R_g, using the Guinier approximation is 3.98 nm, which differs markedly from the *in silico* calculated value using SOMO and the high-resolution model of the corresponding region of Ag43 (1.6 nm) (Figure 6-9B). Estimation of the molecular mass of the particle, based on the I(0) suggests that the protein is monomeric and no obvious aggregation is observed (I(0)= 18.99 kDa), consistent with the AUC experiments (Figure 6-6). The Kratky plot, which provides a qualitative measure of the folding state of the particle, shows a hyperbolic curve with increasing scattering angle, characteristic of a highly flexible, unfolded polypeptide chain (Figure 6-9C). Globular, rigid proteins usually give rise to a bell-shaped curve that declines at higher q values, and deviations from this feature indicate the presence of molecular flexibility and unstructured regions in the protein. The distance distribution plot p(r) of $Ag43_{N168}$ shows a highly asymmetric, broad distribution of electron pairs in the molecule with a long tail at large distances, reaching the value of 17 nm (D_{max}), which indicates a significant deviation from a globular particle (Figure 6-9D).

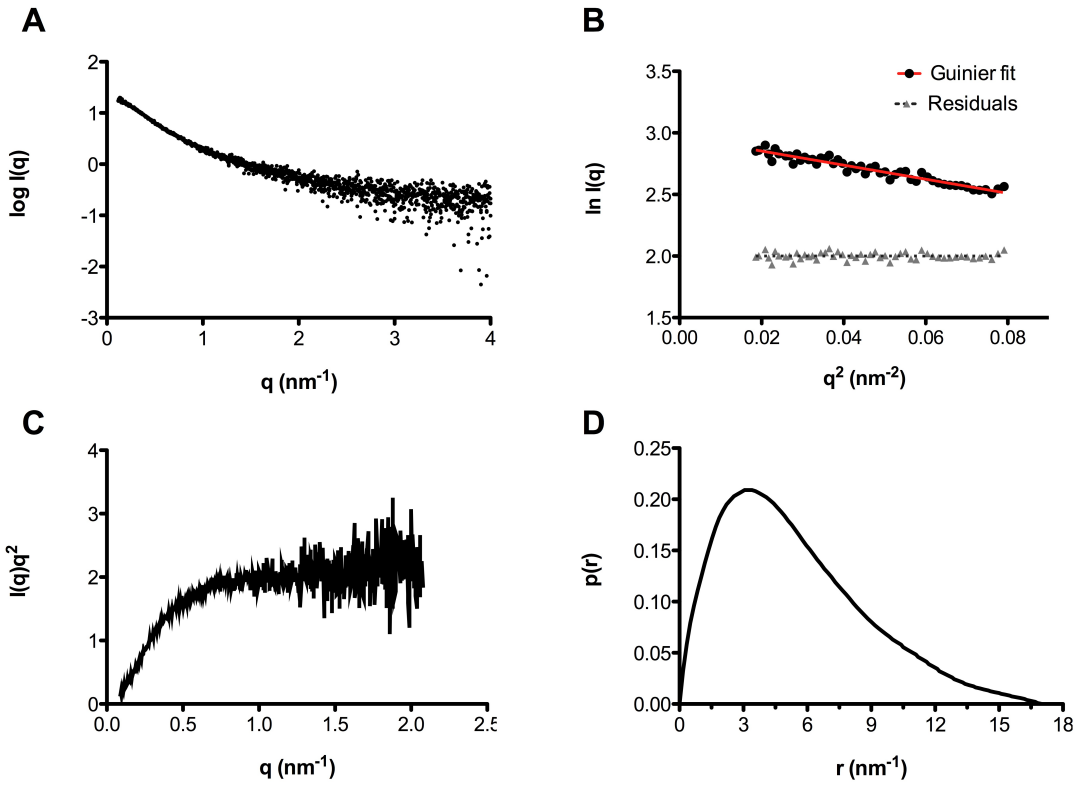


Figure 6-9 Small-angle X-ray scattering analysis of $Ag43_{N168}$ confirms that the molecule exists in a state of intrinsic disorder. (A) Raw scattering curve of $Ag43_{N168}$ showing a featureless shape represents scattering from an average of all solution conformations of the disordered polypeptide. (B) Determination of R_g of from the Guinier region of the scattering curve (R^2= 0.925). Residuals are plotted to show that there is no inter-particle interference in the sample. (C) Kratky plot indicates high flexibility and a lack of globular character in the molecule. (D) Pair-distance distribution plot reveals that the particle is highly elongated in solution with an estimated D_{max} of 17 nm.

The above model-free analysis of SAXS data suggests that $Ag43_{N168}$ exists as a highly dynamic, extended polypeptide chain in solution. However, since this represents the average structure of $Ag43_{N168}$ in solution, the existence of more compact, possibly semi-folded conformers would not be evident from this analysis. Certain IDPs, although devoid of a fixed tertiary structure, can exist as compact structures with some secondary structure elements present (198). However, the majority of known IDPs continuously sample conformational space from compact to extended particles in solution (201). In order to probe the structural ensembles of $Ag43_{N168}$ in solution and to further investigate the extent of molecular disorder, discrete molecular dynamics (DMD) simulations were

performed using US-SOMO on the 3D model of $Ag43_{N168}$ (156). Briefly, the last β-strand of the protein (residues 162-168) was kept static, while the rest of the molecule was subjected to a sequence of atomic collisions which lead to gradual unfolding of the non-static polypeptide, with model "snapshots" taken along the unfolding trajectory. 5000 models were generated in ordered to obtain an evenly distributed random population encompassing compact as well as unfolded states of $Ag43_{N168}$, with several selected examples of the process shown in Figure 6-10.

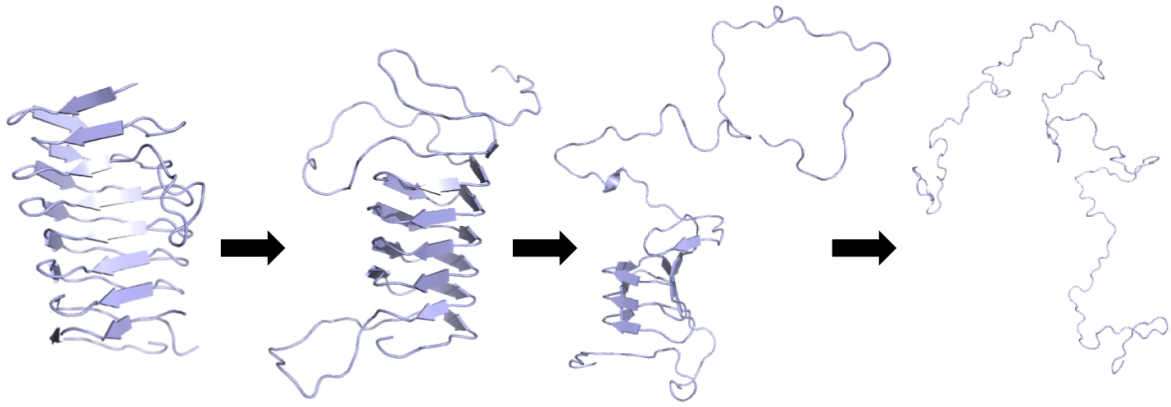


Figure 6-10 DMD simulation of unfolding of $Ag43_{N168}$ using the appropriate region from the 3D model of the crystal structure of Ag43 passenger domain. 5000 models were generated along the denaturing pathway to ensure a Gaussian distribution of the model pool with folded, partially folded and completely unfolded models.

The 5000 generated models were subjected to a genetic algorithm, which is implemented by the program GAJOE in order to select ensembles of these models of $Ag43_{N168}$ that would best describe the SAXS data (155). The genetic algorithm in GAJOE selected models that yielded an excellent fit to the raw scattering data (χ=0.93) (Figure 6-11A). Inspection of the distribution of the selected ensembles suggests that predominantly extended particles best describe the scattering data with a few less extended populations present. The selected distribution is shown in Figure 6-11B, showing that the D_{max} of ensembles ranges from 7.4 – 13.35 nm, with the predominant population at 13.35 nm. A similar distribution of R_g values is also seen (with a range between 1.92-4.5 nm, data not shown) with the majority of molecules with R_g values of 4.4 nm. 3D models of selected

ensembles are shown in Figure 6-11C purely for pictorial representation of average ensemble constituents.

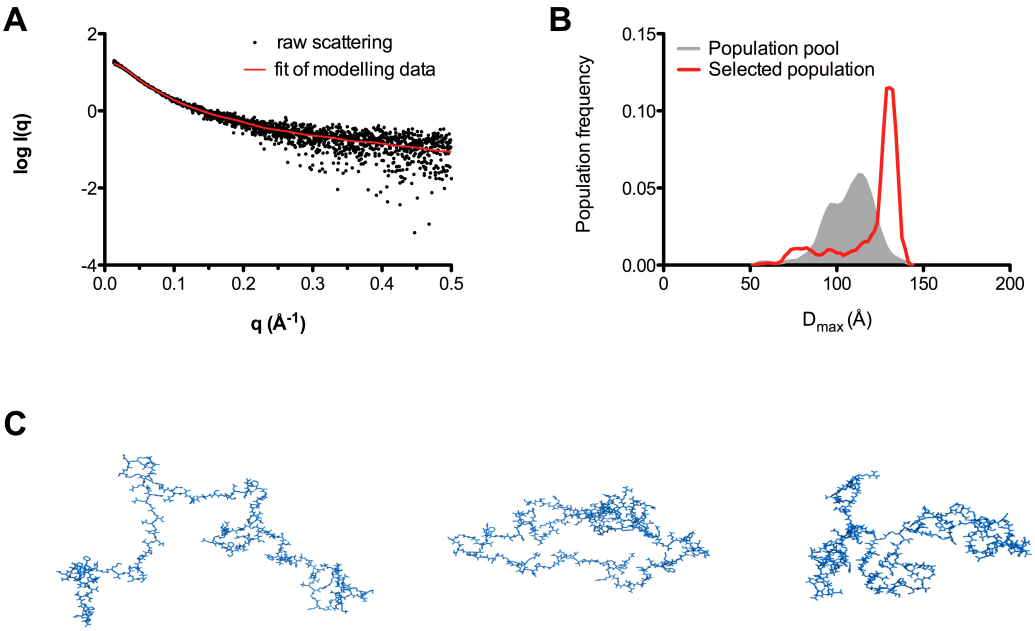


Figure 6-11 Selection of ensembles of Ag43$_{N168}$ against SAXS data using a genetic algorithm. (A) Fit of the selected ensembles of Ag43$_{N168}$ against raw scattering data shows an excellent fit (χ=0.93). (B) Distribution of selected ensembles by their D_{max} and population frequency shows that extended particles dominate the solution conformations with only minor contributions from the more compact species. (C) Representative selected ensembles of extended and compact models.

The finding that Ag43$_{N168}$ exists in a state of disorder is consistent with the previously reported characterisation of the N-terminal region of pertactin, which exhibited a random-coil-like property (133). Interestingly, unlike the majority of IDPs, Ag43$_{N168}$ exists predominantly as an extended polypeptide, rather than sampling the conformational continuum, suggesting that this state of the polypeptide might mimic the structure of the translocating passenger domain prior to its secretion across the OM.

6.2.5 *TamA contributes to the secretion of Ag43 passenger domain*

Folding studies on the pertactin passenger domain revealed that the molecule possesses very slow folding kinetics, on the hours time scale, with a fast-folding C-terminal region and a slow-folding N-terminal region that is dependent on the formation of a stable C-terminal core. This property was attributed to the prevention of the passenger domain folding during its periplasmic life-time in order to maintain it in a translocation-competent state (132). Since the purified $Ag43_{N168}$ could represent such a translocation competent conformation of the passenger domain it forms an ideal substrate to investigate interactions between $Ag43_{N168}$ and TamA and TamB *in vitro*.

Since TamA belongs to the large Omp85 superfamily of protein transporters it could serve as an additional conduit for passenger domain secretion akin to the proposed role of BamA (202). The dependence of Ag43 surface delivery on the presence of TamA was tested using western blotting with antibodies generated against the N-terminal region of Ag43. *E. coli* cells bearing Δ*tamA* were obtained from the Keio collection (146). Additionally, full-length *tamA* including the signal sequence was cloned into a pBad expression plasmid under the control of an L-arabinose-inducible promoter (named pKM55, provided by Dr. K. Mosbahi) and transformed into chemically competent Δ*tamA* *E. coli* cells. Complementation of the knock-out phenotype was tested using a variety of L-arabinose concentrations (0.1-0.005% v/v); however, cells exhibited poor growth in the presence of L-arabinose (data not shown). It was later discovered that the protein expression from the leaky promoter was sufficient for complementation of the knock-out phenotype. Western blots were carried out on whole *E. coli* cells, grown to an OD_{600} of 2. Secreted Ag43 is auto-processed into two main fragments either on the cell surface or in the periplasm during translocation: the extracellular Ag43α (60 kDa) and membrane-bound Ag43β (50 kDa) subunits (203).

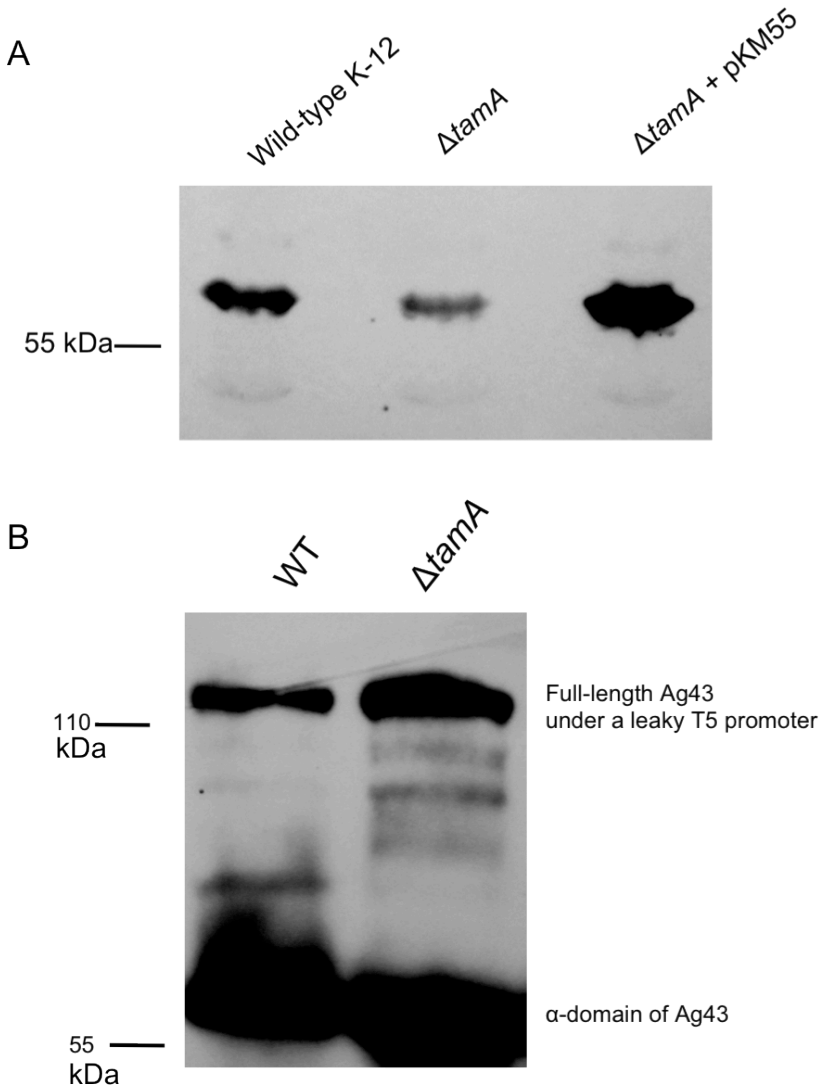


Figure 6-12 Western blots confirm the involvement of TamA in Ag43 secretion. (A) Whole-cell lysates were run on SDS-PAGE gel and blotted using an anti-Ag43 polyclonal antibody. The anti-Ag43 antibody was raised against the purified $Ag43_{N168}$ fragment and therefore detects the presence of the α-domain *in vivo*. pKM55 plasmid bears the full-length *tama* gene for KO complementation. All cells were normalised by OD_{600} to ensure the amount of protein loaded was the same for each sample. (B) Secretion of Ag43 in WT and *ΔtamA* cells expressing Ag43 under a leaky T5 promoter. Cells were grown to an OD_{600} 2, normalised and separated on a gel using SDS PAGE. Accumulation of the full-length Ag43 (at 110 kDa) was most prominent in *ΔtamA* compared with WT.

In the wild-type *E. coli* cells a strong band corresponding to Ag43α was observed (Figure 6-12 A, first lane). Cells lacking *tamA* exhibited a significant reduction in the intensity of the Ag43α fragment, suggesting that little of the processed Ag43 was secreted by the

bacterial cells (Figure 6-12, middle lane). The presence of the band in the Δ*tamA* cells (middle lane) was previously attributed to an endogenous cytoplasmic protein that can cross-react with anti-Ag43α antibody (138). Complementing the knock-out phenotype with full-length TamA under a leaky L-ara-inducible promoter was shown to increase the level of the processed Ag43α, confirming the strong contribution of TamA to the process of Ag43 secretion (Figure 6-12, lane 3).

It was then decided to clone the full-length Ag43 under a leaky T5 promoter (pKM66, provided by Dr Khedidja Mosbahi) in order to attempt and capture the periplasmic intermediates of Ag43. pKM66 was transformed into chemically competent K-12 and *ΔtamA* and blotted. The full-length Ag43 protein was seen to accumulate in the K-12 strain (Figure 6-12 B, first lane). The levels of full-length Ag43 protein in *ΔtamA* were much greater than that of the wild-type cells due to the inability of the protein to be secreted to the bacterial surface (Figure 6-12 B, second lane). This, therefore, suggests that TamA has a direct influence on the secretion of full-length Ag43.

6.2.6 *TamA POTRA domains associate weakly and transiently with the disordered $Ag43_{N168}$ polypeptide*

It has previously been demonstrated that the POTRA domains from BamA weakly associate (mM affinity) with hydrophobic β-strands derived from PhoE, an OMP substrate of the BAM complex (58). Furthermore, the POTRA domains of FhaC were shown to associate with the denatured substrate FHA, suggesting that these domains play a role in substrate detection (97). Therefore, in order to investigate whether this was the case for TamA and Ag43, interactions between $TamA_{POTRA}$ and $Ag43_{N168}$ were tested by NMR spectroscopy. Due to high stability and good spectral dispersion, ^{15}N-labelled $TamA_{POTRA1}$ was used. Since weak interactions were expected, a large 8-fold excess of $Ag43_{N168}$ was titrated into ^{15}N-labelled $TamA_{POTRA1}$. Small, concentration-dependent chemical shifts were observed in a discrete subset of peaks of the HSQC spectra, suggesting that there are weak transient interactions between $TamA_{POTRA1}$ and $Ag43_{N168}$ (Figure 6-13). Chemical shift perturbation data are summarised in Figure 6-13. Due to good chemical shift dispersion of $TamA_{POTRA1}$ cross-peaks, sequential backbone assignment was carried out

using ^{15}N-labelled TamA$_{POTRA1}$ and 3D HSQC-NOESY and 3D HSQC-TOCSY spectra. Approximately 80% of the backbone residues could be assigned with confidence, which allowed the analysis of the binding surface on POTRA1 (Figure 6-14).

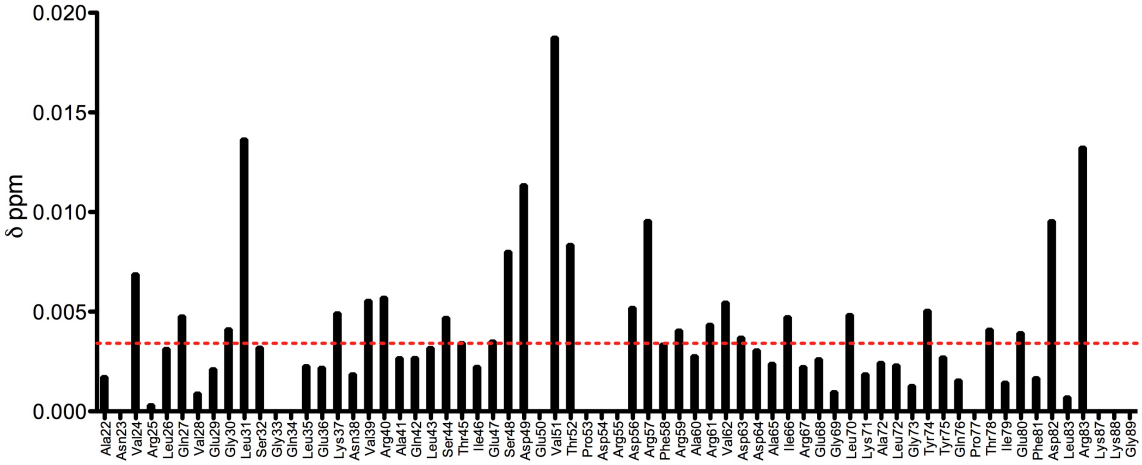


Figure 6-13 Summary of the chemical shift perturbation assay between ^{15}N-labelled TamA$_{POTRA1}$ and unlabelled Ag43$_{N168}$. Shifts within spectra were converted to chemical shift perturbation (CSP) values using the equation $\Delta ppm = \sqrt{[\Delta\delta HN + (\Delta\delta N * \alpha N)^2]}$. Red line represents one standard deviation from the average chemical shift value.

Mapping the chemical shift data onto the high-resolution model of POTRA1 revealed that the putative Ag43 binding site consisted of an extended surface predominantly between the first β-strand and first α-helix of POTRA1. This interaction is likely to be highly electrostatic in nature since chemical shift changes are observed in a number of TamA$_{POTRA1}$ arginine residues.

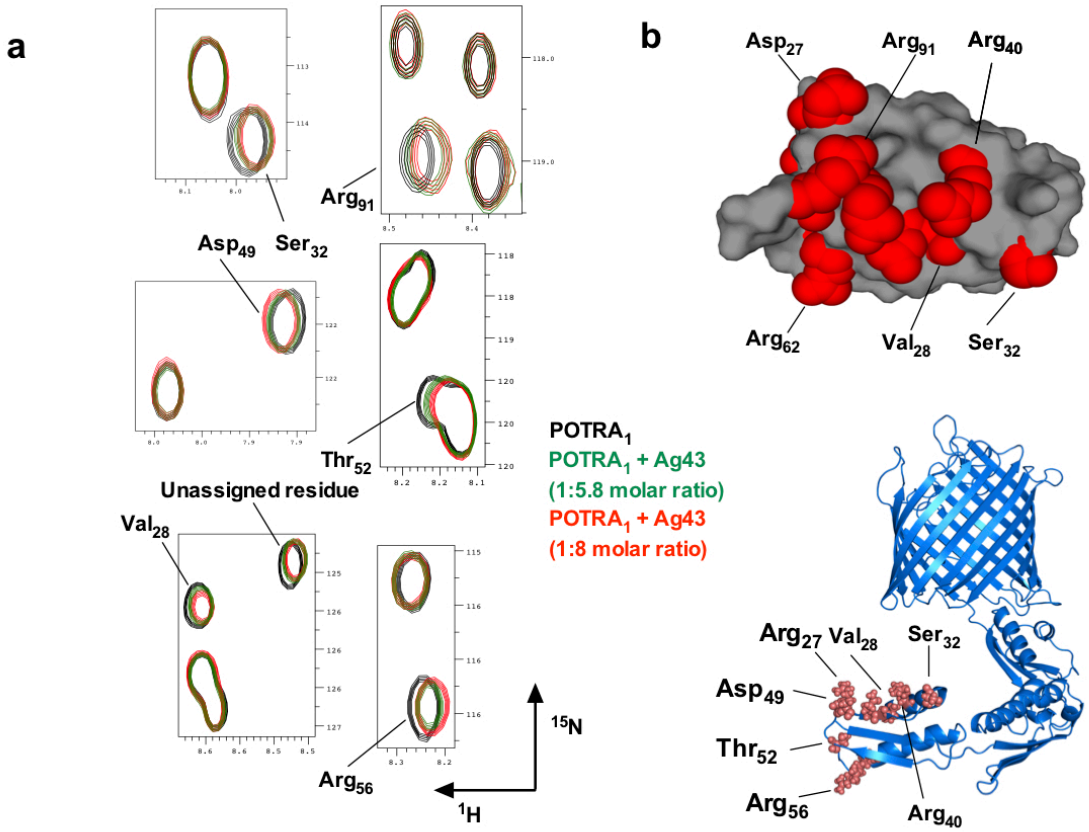


Figure 6-14 Transient electrostatic interactions govern the association between POTRA1 from TamA and the intrinsically disordered Ag43 passenger domain. (A) Small, concentration-dependent chemical shifts can be seen on addition of an 8 molar excess of disordered $Ag43_{N168}$. (B) Binding surface mapped onto POTRA1 (top) as well as full-length TamA (bottom). All binding reactions were carried out in 0.1 M sodium phosphate buffer, pH 6.9.

In order to validate the binding reactions between $TamA_{POTRA1}$ and $Ag43_{N168}$ the reverse titration was carried out, using ^{15}N-labelled $Ag43_{N168}$ and an excess of unlabelled $TamA_{POTRA1}$. In this experiment, no chemical shift changes were observed, however a significant reduction in the intensity of a number of cross-peaks was evident which suggests that $TamA_{POTRA1}$ restricts the local motion of the disordered $Ag43_{N168}$ polypeptide backbone without any detectable effect on the chemical environment of the interacting amino acid residues (Figure 6-15).

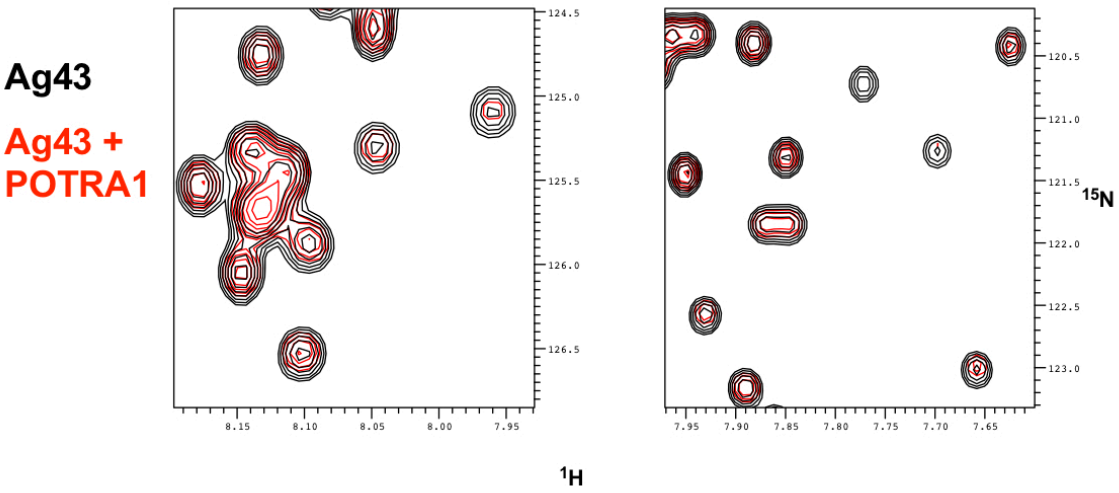


Figure 6-15 Titration of excess $TamA_{POTRA1}$ into ^{15}N-labelled $Ag43_{N168}$ confirms interactions between $TamA_{POTRA}$ and disordered $Ag43_{N168}$. Selected cross-peaks show a reduction in peak intensity upon addition of a 20-fold excess of $TamA_{POTRA1}$ (red peaks, 2 mM) to $Ag43_{N168}$ (black peaks, 100 μM). Spectra were recorded in 50 mM sodium phosphate, 50 mM NaCl, pH 6.9 at 278 K.

6.2.7 *$TamA_{POTRA}$-$Ag43_{N168}$ interactions can be studied using paramagnetic relaxation enhancement (PRE) NMR spectroscopy*

In-depth structural characterisation of $Ag43_{N168}$ - $TamA_{POTRA}$ interactions lies outside the scope of this project. However, some preliminary experiments have been carried out. It was conceived that the most optimal and convenient way to study the interactions between the POTRA domains of TamA and the disordered $Ag43_{N168}$ polypeptide would be to utilise paramagnetic relaxation enhancement (PRE) NMR spectroscopy with ^{15}N-labelled $Ag43_{N168}$ and $TamA_{POTRA}$ domains with paramagnetic spin-labels attached at various sites within the POTRA domains. PRE is an ideal method to study low-population species in fast exchange and is the method of choice for obtaining structural information for weak protein-protein complexes and encounter complexes (204,205). The presence of a delocalised electron within the paramagnetic tag enhances the R_1 and R_2 relaxations of the nuclei at sites in the proximity of the paramagnetic tag in a distance-dependent

fashion. As a test experiment, a cysteine was introduced into $TamA_{POTRA1}$ at position G31 using site-directed mutagenesis and MTSSL ((1-Oxyl-2,2,5,5-tetramethylpyrroline-3-methyl) methanethiosulfonate), a commonly used paramagnetic tag was covalently attached via the formation of a disulfide bond as described in the Material and Methods (section 2.15). Next, ^{15}N-labelled $Ag43_{N168}$ was mixed at a 1:3 ratio with the labelled $TamA_{POTRA1}$ and ^{1}H-^{15}N HSQC spectra were acquired. $Ag43_{N168}$ mixed with buffer served as a control.

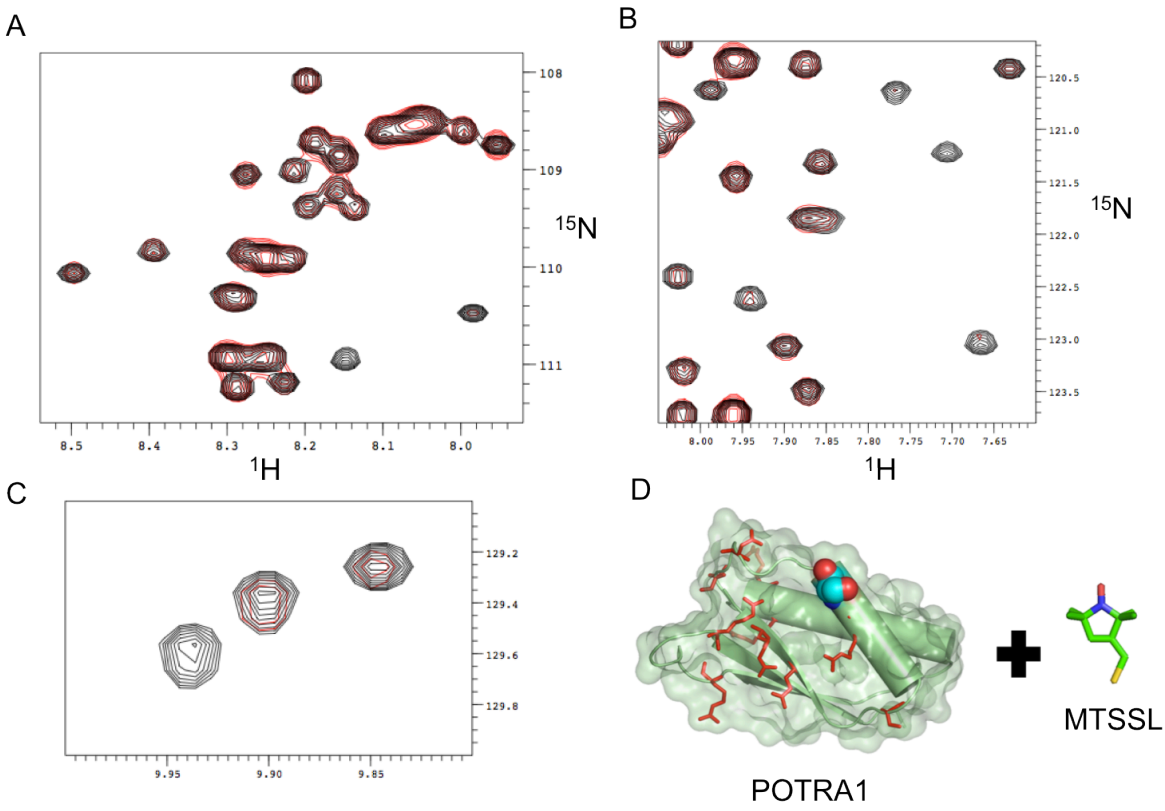


Figure 6-16 Interactions between $TamA_{POTRA1}$ and $Ag43_{N168}$ can be studied by PRE NMR spectroscopy. Selected regions of the $Ag43_{N168}$ HSQC spectrum (A) and (B) reveal peak broadening in ^{15}N-labelled $Ag43_{N168}$ due to the presence of paramagnetically tagged $TamA_{POTRA1}$ domain. (C) The trp-indole group region of the HSQC spectrum implies the involvement of Trp residues in the $Ag43_{N168}$-$TamA_{POTRA1}$ association. (D) Sticks show the putative binding site of $Ag43_{N168}$ on $TamA_{POTRA1}$ reported earlier (section 6.2.6). In cyan is the Gly residue chosen for mutagenesis to Cys for covalent linkage with MTSSL. The buffer used for the experiment was 0.1 M sodium phosphate, pH 6.9.

Selective reduction in the cross-peak intensity was observed in the HSQC spectrum of ^{15}N-Ag43$_{N168}$ + MTSSL-TamA$_{POTRA1}$ compared with ^{15}N-Ag43$_{N168}$ alone; this reduction in peak intensity signified enhanced relaxation and therefore suggested that some specific region on the Ag43$_{N168}$ backbone associates with TamA$_{POTRA1}$ in the proximity of the paramagnetic tag (Figure 6-16).

6.2.8 *Putative role of TamB in the recognition of the unstructured N terminal AT passenger domain region*

The major periplasmic chaperones SurA and Skp have been described as functional chaperones during AT passenger domain passage through the periplasm (113,115). However, inspection of the Ag43 passenger domain sequence reveals no obvious SurA recognition motifs (rich in aromatic residues), which suggests that SurA may not be a major chaperone for Ag43 passenger domain biogenesis. The formation of a cell envelope-spanning complex between TamA and TamB suggests that TamB may in fact act as a membrane-bound chaperone for the delivery of AT passenger domains to TamA, playing a role analogous to the soluble chaperones Skp and SurA in the delivery of integral membrane β-barrel domains to the BAM complex.

The interactions between the refolded TamB$_{DUF490}$ construct and Ag43$_{N168}$ were probed using analytical size-exclusion chromatography using a Superdex S75 GL 10/300 column. An excess of TamB$_{DUF490}$ was mixed with Ag43$_{N168}$, allowed to equilibrate and applied onto the column. Unexpectedly, a shift was observed in the elution profile of TamB$_{DUF490}$ in the presence of Ag43$_{N168}$, compared with the two proteins alone, which suggests that the two molecules are forming a complex, present in the leading boundary of eluting TamB$_{DUF490}$ (Figure 6-17A).

Ag43$_{N168}$-SurA interactions were also probed using analytical SEC. Protein solution consisting of a 1:2 SurA:Ag43$_{N168}$ mixture of both proteins was applied to an analytical Superdex S75 column, with individual proteins run down the column as a control. No change in the elution boundary or peak position was observed, with the mixture of proteins eluting at the same positions as the proteins alone, suggesting no detectable

binding takes place between the two molecules (Figure 6-17B). Complex formation between TamB and $Ag43_{N168}$ was investigated in further detail using biophysical and structural methods.

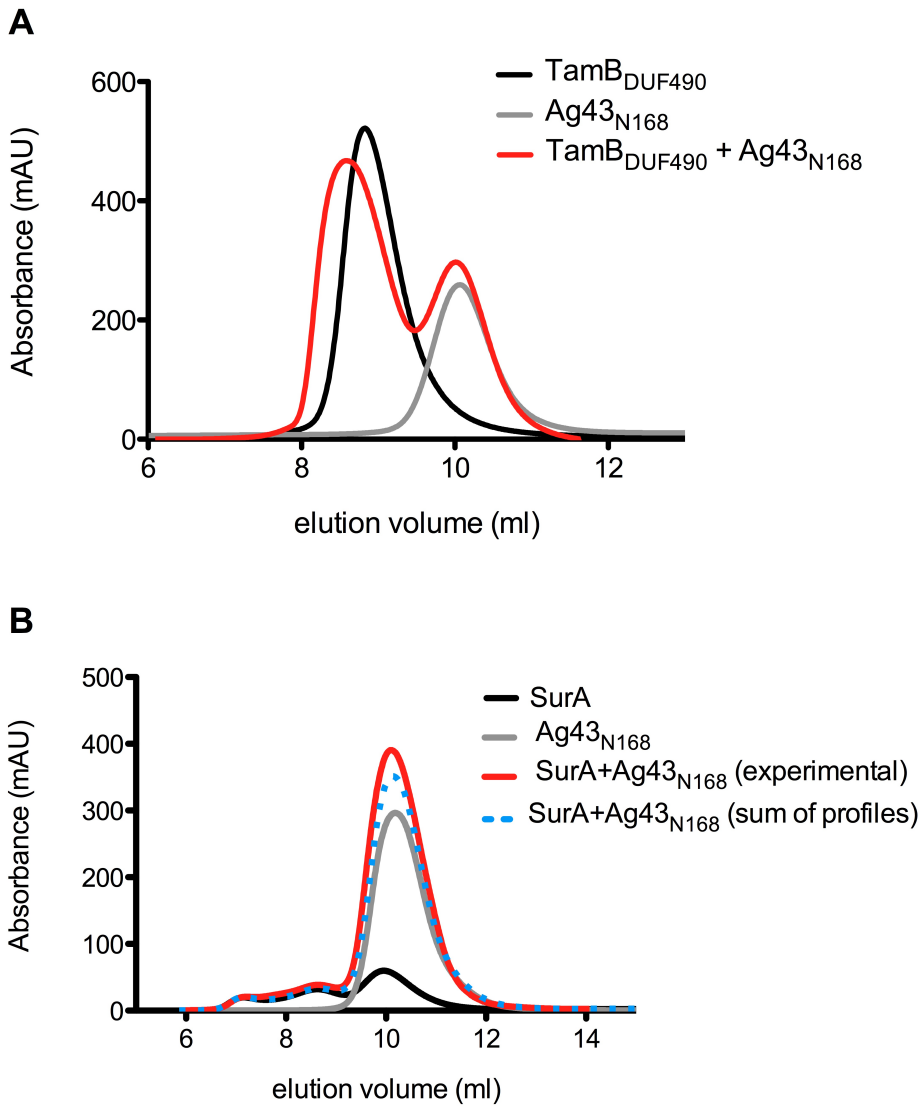


Figure 6-17 Assaying for interactions between $Ag43_{N168}$ and SurA (A) and $Ag43_{N168}$ and $TamB_{DUF490}$ (B) by analytical SEC. The column used in this experiment was an analytical Superdex S75. SurA (66 μM), Ag43 (150 μM) and both proteins with equivalent concentrations (approximately 1:2 stoichiometric mixture) were made up and allowed to reach equilibrium for 15 minutes at room temperature prior to injection onto the column. SurA exhibited some multimeric character and previously it has been shown to associate to substrate peptides as a dimer (45), however higher M_w species could also be formed. The buffer selected included 0.3% CHAPS detergent to promote the stability of $TamB_{DUF490}$ owing to its small micelle size. An excess of $TamB_{DUF490}$ (200 μM) was mixed with $Ag43_{N168}$ (100 μM), allowed to equilibrate and injected onto the column.

ITC experiments with $TamB_{DUF490}$ and $Ag43_{N168}$ indicate enthalpy driven association between the two proteins with a K_d of 5.6 ± 0.8 μM (Figure 6-18). Additionally, the interactions between the putative periplasmic region of TamB, namely $TamB_{C80}$, and the $Ag43_{N168}$ construct were probed. Similar $Ag43_{N168}$ binding was observed with $TamB_{C80}$ and $TamB_{DUF490}$, suggesting that the C-terminal region of TamB is also involved in Ag43 recognition, in addition to TamA association (Figure 6-19 A).

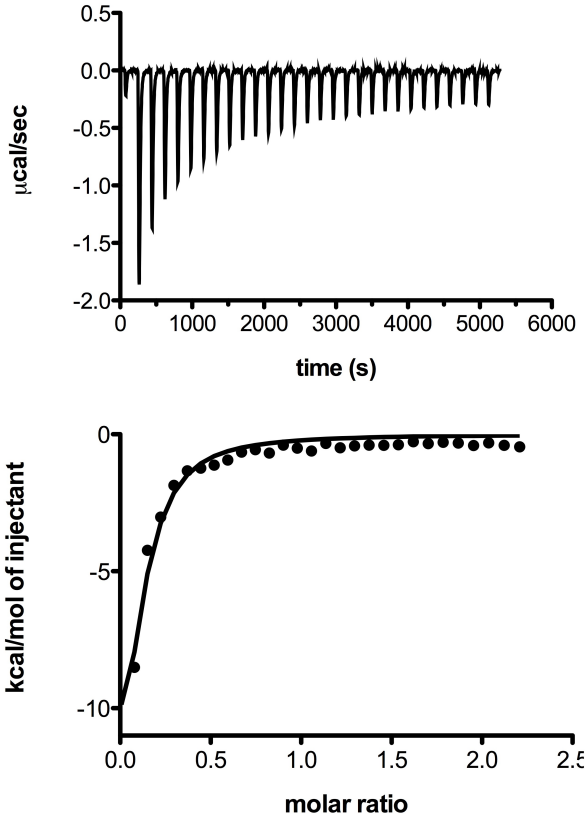


Figure 6-18 Binding between $TamB_{DUF490}$ and $Ag43_{N168}$ detected by ITC. Raw thermogram of the binding reaction (top), and the binding isotherm with the integrated heat injections fitted with a single-site binding model (bottom). The binding reactions were carried out in 0.2 M sodium phosphate, pH 7.5, 0.015% Triton-X-100. 500 μM $TamB_{DUF490}$ was titrated into 50 μM $Ag43_{N168}$.

Since $TamB_{C80}$ exhibited much higher stability than $TamB_{DUF490}$ a higher concentration of $TamB_{C80}$ was used in the binding assay. It uncovered a putative second binding site on $Ag43_{N168}$ because single-site and two-site binding models could not be fitted with the integrated heats of the isotherm (Figure 6-19 B). Instead a sequential two-site model

produced a good fit and yielded two different dissociation constants, K_{d1} of 4.5 μM and K_{d2} of 500 μM. Interestingly, the thermodynamic profiles of the two reactions were different. For site 1 (K_{d1}) the binding exhibited favourable enthalpic and entropic contributions (negative ΔH and positive TΔS), whereas for site 2 the binding was enthalpically driven with unfavourable entropy (negative ΔH, negative TΔS). This suggests that $Ag43_{N168}$ possesses two distinct binding sites for TamB, a high affinity as well as a low affinity site. Attempts to investigate TamB-$Ag43_{N168}$ interactions using stopped-flow fluorimetry were unsuccessful since no change in fluorescence was observed when the two proteins were mixed together (data not shown).

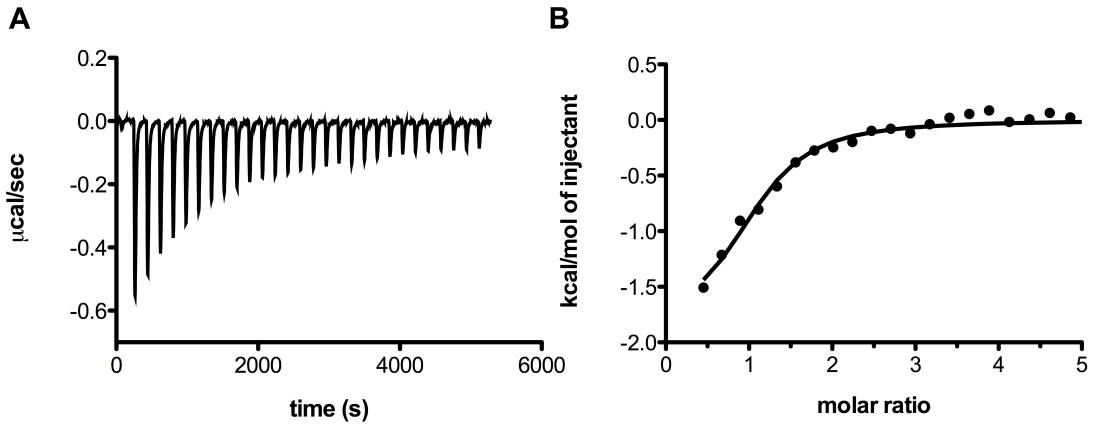


Figure 6-19 Interactions between Ag43N168 and TamBC80 indicate that the C-terminal region of TamB is responsible for substrate association. (A) Raw ITC data showing saturable binding by titrating 2 mM TamBC80 into 50 μM Ag43N168. (B) Integrated heats of binding produce an isotherm with a measured Kd of 3 μM. Data were fitted with a sequential two-site binding site model, after single-site binding and two-site binding models failed to produce good fits to the data. ITC was performed in 0.2 M sodium phosphate buffer, pH 7.5, 0.015% Triton X-100, 25°C.

6.2.9 *$TamB_{DUF490}$-$Ag43_{N168}$ interactions can be monitored using chemical shift perturbation NMR spectroscopy*

TamB-$Ag43_{N168}$ interactions were next probed by ^{1}H-^{15}N HSQC NMR chemical shift perturbation assay using ^{15}N-labelled $Ag43_{N168}$ and $TamB_{C80}$ to obtain some structural details about the association between the two proteins. Addition of $TamB_{C80}$ to ^{15}N-labelled $Ag43_{N168}$ resulted in chemical shift changes as well as changes in peak intensity in

a large number of cross-peaks (Figure 6-20). At least 30% of the cross-peaks were affected by $TamB_{C80}$, yet the peak dispersion of the $Ag43_{N168}$-$TamB_{C80}$ complex does not correlate with folding-upon-binding, since the ^{1}H cross-peak dispersion was not expanded. These data are suggestive of global changes to the backbone conformation of the Ag43 passenger domain on binding of TamB, which indicates that TamB could be associating with a specific linear region of $Ag43_{N168}$ and maintaining the protein in a state of disorder, or a translocation-competent state. This finding could have functional significance for passenger domain translocation though the TamA barrel, wherein TamB imposes a conformation on the Ag43 backbone which favours its translocation by preventing premature folding or translocation of the N-terminal region of the passenger domain. Since TamB can associate with TamA and $Ag43_{N168}$ it can therefore act as a periplasmic chaperone to divert the periplasmic passenger domain towards TamA.

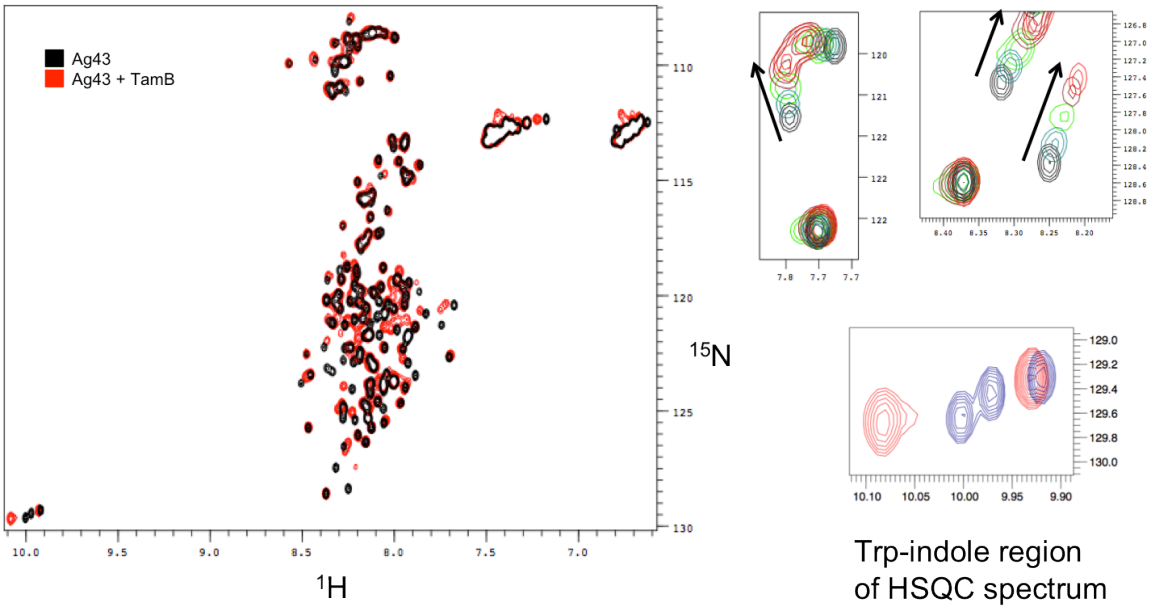


Figure 6-20 ^{1}H-^{15}N HSQC spectrum of $Ag43_{N168}$ in the presence (red) and absence (black) of excess $TamB_{C80}$. Selected concentration-dependent chemical shifts are plotted to show that resonance perturbation is concentration-dependent. Additionally, changes in chemical shifts of Trp residues are shown by inspecting the indole resonance region of the spectrum. 100 μM of ^{15}N-labelled $Ag43_{N168}$ and 1 mM $TamB_{C80}$ were used for the binding reaction. Spectra were acquired in 0.1 M sodium phosphate, pH 6.9, 0.015% Triton X-100.

Sequence-specific backbone assignment of $Ag43_{N168}$ was attempted using 3D HSQC-NOESY and 3D HSQC-TOCSY spectra; however, only 32 of 169 resonances could be assigned with confidence. Complete assignment of the $Ag43_{N168}$ backbone would require the acquisition of additional spectra such as HNN and HN(C)N which are commonly used for backbone assignment of IDPs on the doubly labelled ^{15}N, ^{13}C $Ag43_{N168}$ lies outside the scope of this study (206).

From the inspection of the ^{1}H-^{15}N HSQC spectra at least 2 of the 3 Trp residues are affected by the interaction with $TamB_{C80}$, and could thus serve as putative regions of hydrophobic interaction with $TamB_{C80}$ (Figure 6-20). This finding raises the possibility that there are several binding sites within the unfolded polypeptide, a high affinity site, as well as one or more putative weak binding sites since the Trp residues are located far apart in the amino acid sequence of $Ag43_{N168}$ (Figure 6-21). This is further supported by the ITC data involving $TamB_{C80,}$ which fit poorly with a single-site binding model. Moreover, there could exist a potential allostery between the N- and C-terminal polypeptide regions in $Ag43_{N168}$ that are recognised by TamB. Incomplete assignment of the $Ag43_{N168}$ backbone prevents formulation of models for the TamB molecular recognition events. Since chemical shift perturbation does not directly reflect the amino acids involved in the intermolecular interactions, but rather the residues that are affected by the associating partner protein overall, additional factors such as changes in peak intensity must be considered, especially for the case of an IDP-like protein such as $Ag43_{N168}$.

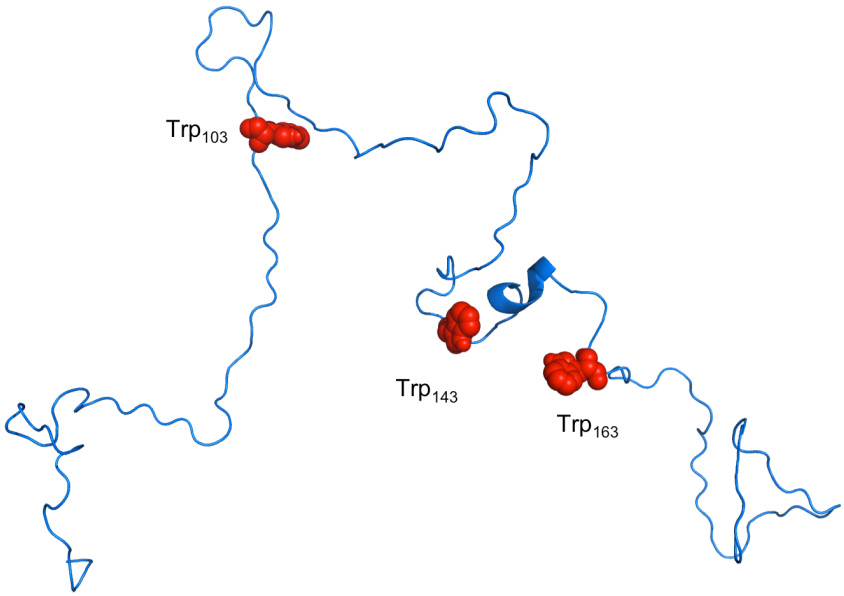


Figure 6-21 Positions of the three Trp residues highlighted in the putative extended conformation of $Ag43_{N168}$ proposed from SAXS analyses. Two of the three highlighted Trp residues are affected by the association with $TamB_{C80.}$

6.2.10 *TamA and TamB function in tandem to aid autotransporter secretion*

Since the same TamB domain was involved in TAM complex formation with TamA as well as substrate association, it was speculated that there might be a possibility of binding site proximity or overlap between TamA and Ag43. Therefore, the interactions between $Ag43_{N168}$ and several truncated TamB proteins were tested to see whether the binding site overlaps with or is close to that of TamA. Interestingly, no detectable heats of binding were observed upon titration of $TamB_{DUF490\Delta35}$ and only very weak heats could be detected with $TamB_{DUF490\Delta20}$. This suggests that the C-terminus of the protein has a bi-functional role in substrate recognition as well as TamA association, since the binding between $TamB_{DUF490\Delta20}$ and $TamA_{POTRA}$ was not significantly perturbed (Figure 5-9). Moreover, this finding is also indicative of a possible binding site overlap between TamA and substrate passenger domains, which may suggest a sequential mechanism of substrate recognition and TamA association.

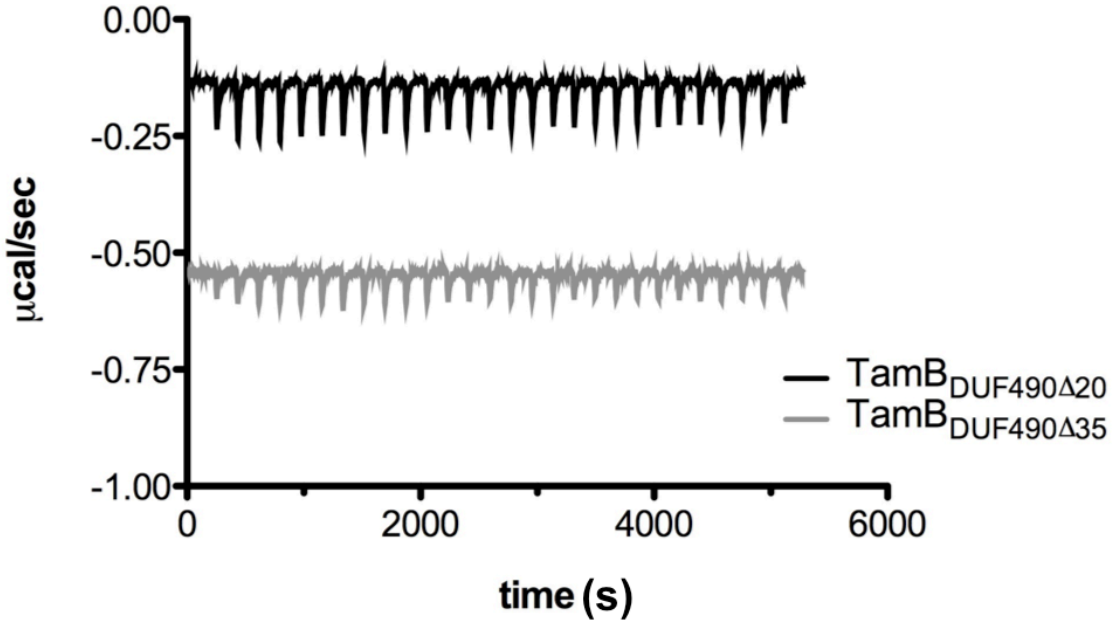


Figure 6-22 Probing interactions between $Ag43_{N168}$ and $TamB_{DUF490}$ truncation mutants. All binding reactions were performed in 0.2M sodium phosphate pH 7.5 + 0.015% Triton X-100. 50 μM $Ag43_{N168}$ was placed in the ITC cell and 300 μM $TamB_{DUF490}$ mutants in the syringe. No binding was detected with $TamB_{DUF490\text{-}35}$ or with $TamB_{DUF490\text{-}20}$, a construct which was able to bind to $TamA_{POTRA}$ domains. This analysis is suggestive of the existence of adjacent TamA and Ag43 binding sites within $TamB_{DUF490}$ towards.

Attempts were made to study the interaction using $Ag43_{N168}$ and the pre-formed $TamA_{POTRA}$-$TamB_{DUF490}$ complex using ITC, however numerous attempts were hampered by technical issues. At high concentrations and in the presence of any detergent (LDAO, Triton X-100 and CHAPS) $Ag43_{N168}$ forms a solid gel-like substance (possibly due to amyloid formation), which might be related to its function as an auto-aggregative protein *in vivo*.

However, one attempt at the competition assay was made using NMR spectrometry. The complex between ^{15}N-labelled $TamA_{POTRA12}$ (60 μM) and unlabelled $TamB_{DUF490}$ (150 μM) was made, and the HSQC spectra exhibited the expected broadening of the peaks associated with POTRA1 domain (Figure 6-23 A). Upon addition of subsaturating amounts of $Ag43_{N168}$ (80 μM) some recovery of the broadened POTRA1 peak resonances was seen, suggesting that $Ag43_{N168}$ is able to outcompete the bound $TamB_{DUF490}$ from the $TamA_{POTRA12}$ (Figure 6-23 B). Due to a limited amount of $Ag43_{N168}$ sample this experiment

was not continued further. Nevertheless, it can be concluded that there exists competition between TamA and Ag43 for TamB, implying that the function of TamA and TamB are concerted and sequential rather than together as a complex.

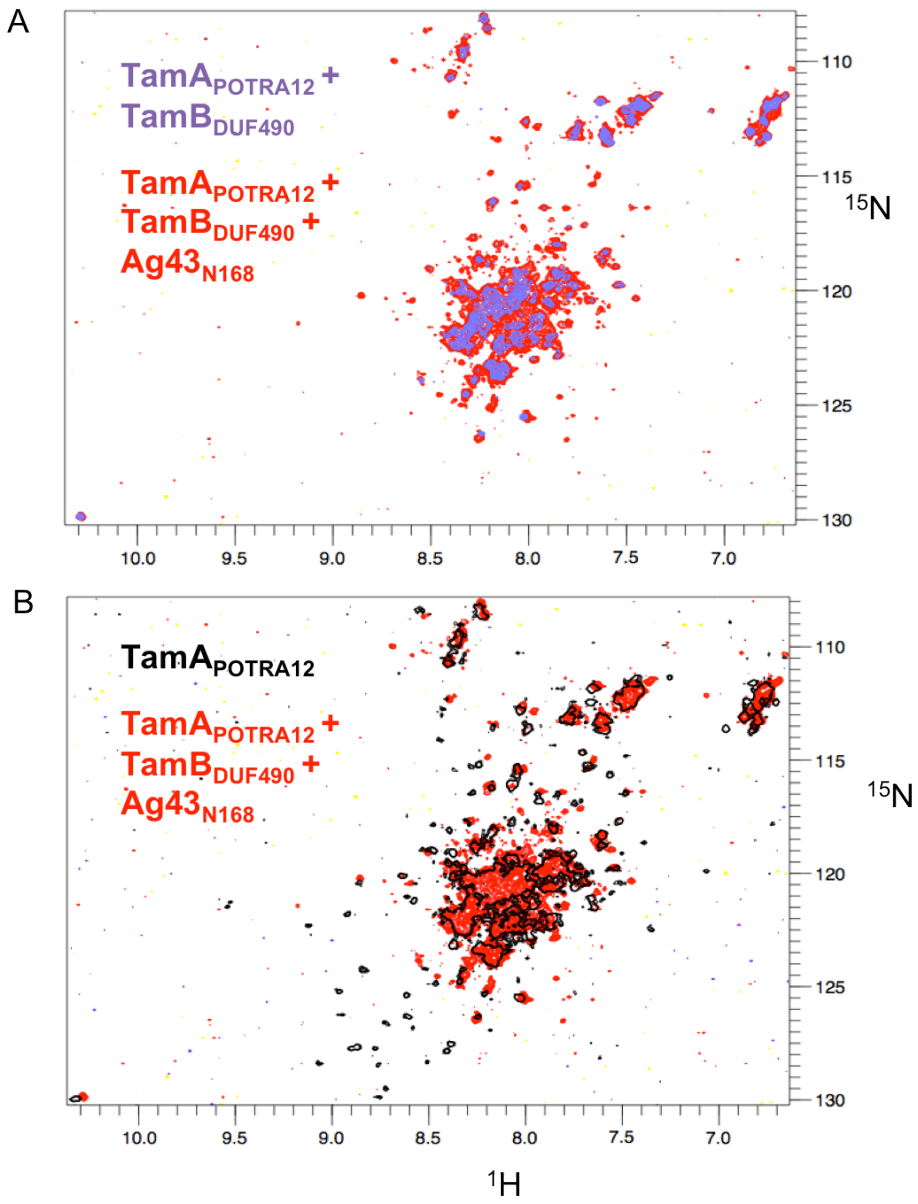


Figure 6-23 Competition binding between TamA and Ag43$_{N168}$ suggests a sequential mechanism. (A) ^{1}H-^{15}N HSQC spectrum of the pre-formed TamA$_{POTRA12}$ - TamB$_{DUF490}$ complex. Poor spectral quality is attributed to the presence of Triton X-100 detergent and low sample quantity. (B) The pre-formed TAM complex is outcompeted by the addition of sub-saturating amounts of unlabelled Ag43$_{N168}$. The outcompetition is characterised by the appearance of resonances arising from the broadened cross-peaks associated with POTRA1. Buffer used in the experiments was 50 mM sodium phosphate, 100 mM NaCl, pH 6.9, 0.015% Triton X-100.

6.3 Discussion

AT biogenesis is a topic of great interest due to its direct link to bacterial pathogenesis and is a possible target for therapeutic intervention, therefore it is crucial to understand it in molecular detail. Several periplasmic chaperones are involved in the biogenesis of ATs during their secretion across the OM (116). The TAM complex, consisting of TamA and TamB, has recently been implicated in aiding the secretion of AT passenger domains in Gram-negative bacteria (138). In this chapter, the interactions between $Ag43_{N168}$ and inidividual TamA and TamB proteins were characterised. TamB has the capacity to interact with an unstructured N-terminal region of Ag43 passenger domain and displays an affinity for $Ag43_{N168}$ similar to its affinity for the POTRA domains of TamA. TamA, on the other hand can associate with the unstructured $Ag43_{N168}$ weakly and transiently. The finding that the N-terminal region of $Ag43_{N168}$ behaves as a random-coil-like polypeptide is in accord with previous studies that probed the folding and stability of AT passenger domains (127,131). A common feature is apparent wherein the C-terminal region of the passenger domain exhibits highest stability and faster folding kinetics and is thought to form the scaffold for the rest of the passenger domain folding (132). The N-terminus, on the other hand, has been shown to exist in a state of random-coil in the absence of the C-terminal stable region (127). The functional implications of this interaction are discussed below.

ATs cross the IM and enter the periplasmic space via the Sec system in a fashion that is entirely different from their translocation across the OM. Sec-dependent transport occurs in a N-to-C terminal fashion following ribosomal translation of the emerging polypeptide. Yet studies have shown that passenger domain secretion across the OM begins with the C-terminal region (131). This vectorial translocation is hypothesised to drive the process in the absence of ATP or any other source of external energy. The fast-folding kinetics of the C-terminus as it emerges outside the cell could act as a potential mechanism to prevent back-sliding through the translocator pore (132). Additionally, it could serve as a nucleation site for the folding of the rest of the passenger domain as it emerges across the OM, contributing further as a drive for energy distribution across the OM, with the folding polypeptide on the surface driving the translocation of the unfolded

periplasmic passenger domain (133). Numerous periplasmic chaperones have been shown to contribute to AT secretion, in particular by associating with the passenger domain during its periplasmic life-time (60,115). However, it is unclear whether this is the case for all ATs. Previously characterised interactions between EspP and SurA, FkpA and DegP show that these chaperones associate with the unfolded passenger domain *in vitro* (115,207). These interactions were shown to be dependent on the presence of several aromatic residues and a specific motif, namely, Aro-X-Aro. In this chapter, an additional chaperone, namely TamB, involved in AT passenger domain biogenesis is described. The preliminary binding interactions studied in this chapter reveal that SurA does not detectably associate with $Ag43_{N168}$. Analysis of the amino acid sequence of $Ag43_{N168}$ indicates a low frequency of aromatic residues (4 Tyr, 2 Phe and 3 Trp), all well-separated in the primary structure. Additionally, only 4 Pro residues are found in this polypeptide segment. Therefore, the lack of any interactions between SurA and $Ag43_{N168}$ is not surprising. On the other hand, TamB is able to interact with the N-terminal, unstructured region of the $Ag43_{N168}$ passenger domain with a low μM affinity in a way that does not induce any polypeptide folding in the IDP-like protein region as shown by ^{1}H-^{15}N HSQC binding experiments. These findings implicate TamB in the early stages of periplasmic chaperoning of AT passenger domains since the N-terminal segment is the first region to enter the periplasm. Since the loss of unstructured character of N-terminal passenger domains has been shown to stall their translocation across the OM, it could be speculated that the role of TamB is to keep the N-terminal AT polypeptide in an unfolded, translocation-competent state, possibly shielded from periplasmic chaperones, since it is that region of the passenger domain that crosses the OM last and possesses slow-folding kinetics (133). This idea is further strengthened by the kinetic folding studies of pertactin which have clearly shown the existence of off-pathway folding intermediates within the N-terminal region of the protein, that hamper and slow down the proper folding of the passenger domain (132). There is a possibility that TamB could associate with additional sites on the Ag43 passenger domain, as evident from the ITC data with $TamBC_{80}$. Since only one Ag43 construct was tested, any additional sites of binding across the entire passenger domain have not been examined.

Weak interactions observed between $Ag43_{N168}$ and $TamA_{POTRA1}$ are in accord with previous binding characterisation of BamA POTRA domains and peptides derived from a substrate OMP (58). On the contrary, the interactions measured between the two POTRA domains from FhaC and the N-terminal region of FHA were reported to be in the low μM range (94). This difference could be attributed to the fact that FhaC possesses only a single substrate molecule, whereas the range of substrates for BamA and TamA proteins is most likely quite large. The increased number of substrates, as well as the need for rapid and transient substrate turnover during polypeptide translocation, would require the rapid association and dissociation of substrates with little need for a high affinity towards specific sequence elements. This is especially likely to be true, if the role of substrate selectivity and specificity resides primarily with other members of the transport complex, e.g. TamB in the case of the TAM complex, and BamD and BamB in the case of the BAM complex.

6.4 Conclusions

An N-terminal region of Ag43 passenger domain, residues 54-220 ($Ag43_{N168}$), was recombinantly expressed and purified to homogeneity. Structural and biophysical characterisation of this polypeptide revealed that $Ag43_{N168}$ exists in an IDP-like state. This polypeptide region is able to interact with the C-terminal region of TamB, in a chaperone-like manner, without inducing any folding-upon-binding in $Ag43_{N168}$ as revealed by ITC and NMR spectroscopy. A possible binding site overlap exists between TamA and $Ag43_{N168}$ on $TamB_{C80}$ implying a concerted mechanism of substrate binding and translocation. Sequence-specific backbone assignment of $Ag43_{N168}$ will aid in determining the TamB binding motifs in this polypeptide, which will further our understanding of the mechanism underlying passenger domain recognition

7 Concluding remarks

Efficient secretion of ATs is reliant on the passage of a newly formed polypeptide chain across two membranous bilayers surrounding the bacterial cell. Translocation across the inner bacterial membrane occurs via the general Sec translocon through the recognition of the N-terminal signal sequences found in all ATs (108). Once in the periplasmic space, omnipresent chaperones such as SurA, Skp, DegP and the BAM complex orchestrate efficient secretion of ATs across the bacterial OM (60,112,114,116). Recently, an additional component involved in AT biogenesis was identified, namely the TAM complex (138). This thesis deals primarily with the *modus operandi* of the TAM complex and its putative role in facilitating the translocation of AT passenger domains across the OM.

TAM complex formation has been shown to involve the POTRA domains of TamA and the C-terminal 80 amino acids of TamB (138). In this thesis, TamA-TamB interactions were characterised further using ITC, stopped-flow fluorimetry and NMR spectroscopy. Through domain deletion experiments it was concluded that POTRA1 was dispensable for association with TamB with POTRA2 forming the predominant binding site and POTRA3 providing a minor contribution to complex formation. This finding supports recent investigations into the sequence diversity and functional specialisation amongst the POTRA domains of Omp85 proteins, where the N-terminal POTRA1 and POTRA2 of TamA form a separate cluster based on their protein sequence similarity with other Omp85 members and POTRA3 being most closely related to BamA POTRA domains (208).

Structural analysis of TamA POTRA domains using biophysical and structural techniques has revealed that these domains possess a high degree of flexibility, which is partially attributed to inter-domain mobility. Despite extensive studies into the structure and function of POTRA domains from other molecular systems such as BamA and FhaC, the conformational flexibility within these domains is currently not understood (95,105). However, characterisation of TamA-TamB interactions presented in this thesis using stopped-flow and NMR spectroscopy has revealed that POTRA-associated partner proteins, in this instance TamB, are able to modulate POTRA inter-domain conformations, suggesting that domain flexibility is somehow linked to domain function. It is has been suggested previously that changes in inter-domain orientation of BamA POTRA domains

could assist in the formation of initial β-hairpins during OMP translocation and insertion into the OM (66). Whether this is the case remains to be validated experimentally.

Unlike its partner protein TamA, the role of TamB in AT biogenesis has, to date, been elusive. Absence of sequence homology with any known functional protein, together with little available biochemical or structural information available for TamB make speculations about its function difficult. In this thesis it was established that TamB interacts with both TamA and the N-terminal region of the passenger domain of a substrate AT, Ag43, through its conserved C-terminal DUF490 domain. Preliminary biophysical and structural characterisation of the N-terminal region of DUF490 revealed an extended, flexible particle rich in β-sheets. The availability of well-diffracting crystals of a sub-region of DUF490 (residues 963-1138), generated as part of this project, paves the way for a detailed structural characterisation of this domain, and the availability of a high-resolution structure would then provide some functional insight about this domain. In stark contrast, the C-terminal region of the DUF490 domain has been shown to lack any well-defined, tertiary structure and seemingly possesses bifunctional roles in mediating TAM complex formation and substrate recognition. Interestingly, TamB displays a similar affinity for TamA POTRA domains and $Ag43_{N168}$, and in both cases the interacting region of TamB is located within the C-terminal 35 amino acids, with binding sites that are adjacent and partially overlapping, but are not identical. The lack of well-defined tertiary structure within the TamB C-terminus might allow it to function as a protein recognition platform, that can associate with TamA, as well as recognise a variety of substrate molecules through putative recognition of linear sequence motifs within unstructured AT passenger domains.

The passenger domain of most ATs forms an extended β-helical moiety, which is thought to drive its secretion across the OM (133). *In vitro* folding studies of the pertactin passenger domain revealed a two-step refolding pathway wherein the C-terminus of the passenger domain exhibits fast-folding kinetics and forms a stable core, which is thought to serve as a template for the folding of the rest of the passenger domain in a directional manner as the polypeptide is translocated outside of the bacterial cell (127,132,133). Interestingly, the N-terminus of the pertactin passenger domain is unable to acquire its

native β-helical fold in the absence of the C-terminal stable core and has been shown to exist in a random coil-like state (133). It is not known whether this feature is pertinent to other AT passenger domains. However, structural characterisation of the N-terminal region of Ag43 (residues 54-220) described in this thesis suggests that it also exhibits poor folding propensity and instead exists as an extended, unstructured polypeptide in a state of intrinsic disorder. However, unlike most IDPs, this N-terminal region exists as a predominantly extended molecule, rather than a polypeptide that constantly samples a plethora of conformational states from compact to extended structures. It could be speculated that the extended, unstructured N-terminus of $Ag43_{N168}$ represents a translocation-competent state of the passenger domain, which further supports the hypothesis that TamB acts as a chaperone in the delivery of secreting AT passenger domains to TamA.

AT biogenesis is seemingly reliant on both the BAM and TAM complexes, with the former required for OM insertion of the β-barrel domain and the latter required for efficient secretion of the AT passenger (138,202). Similarly, in addition to binding TamB, which is mediated largely by POTRA2, TamA also interacts, albeit weakly, with the Ag43 passenger domain through POTRA 1 and perhaps additional POTRA domains. These interactions may be largely electrostatic in nature since the TamA POTRA domains are rich in surface arginine and lysine side chains and have a predicted pI of 8.9, while the Ag43 passenger domain has a predominance of negatively charged amino acids and a predicted pI of 5.6. Interestingly, it has recently been shown that reducing the negative charge of the EspP passenger domain impairs its secretion (209). The weak interaction between POTRA domains and substrate molecules is also observed for the BamA POTRA domains and suggests poor target specificity and recognition, indicating that additional components of the BAM and TAM complexes may be devoted to that function (58). In the case of BamA the soluble chaperone SurA interacts directly with its POTRA domains and this is thought to be a mechanism for targeting unfolded polypeptides to the BAM complex (116). For the TAM complex, the binding data presented in this thesis indicate that TamB likely fulfils the role of the chaperone that would be required for delivery of unfolded AT passenger domains to the POTRA domains of TamA, since the binding between the unstructured Ag43 N-terminus and SurA could not be demonstrated. At the present time

it is not known how BamA and TamA cooperate in AT biogenesis but one possibility is that their action is coordinated, with targeting to the OM accomplished by insertion of the AT barrel into the OM by the BAM complex and translocation of the passenger domain across the OM mediated by the TAM complex.

Taken together, the work presented in this thesis suggests a mechanism of AT passenger domain secretion in which TamB initially acts as a chaperone, binding the nascent passenger domain in the periplasm. Subsequent release of the AT passenger domain from this complex is facilitated by competition for binding of TamB with POTRA2 of TamA. The formation of the TamA-TamB complex therefore serves to release the AT passenger domain close to the TamA POTRA domains and additionally reduces the flexibility of the POTRA domains, perhaps locking POTRA1 in a favourable conformation to bind the AT passenger domain and mediating its passage into the lumen of the TamA barrel for translocation across the OM. This putative mechanism is summarised in Figure 7-1.

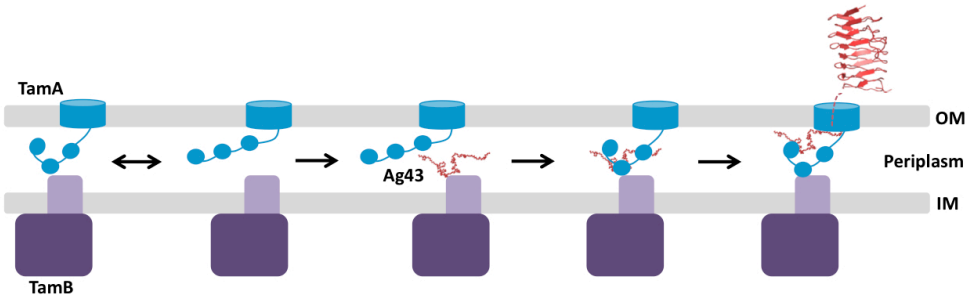


Figure 7-1 Putative mechanism of TAM mediated AT passenger domain secretion. On translocation into the periplasm the AT passenger domain (red) is bound by TamB (purple), which exists in equilibrium between free and TamA (blue) bound forms. TamB delivers the unfolded substrate to the POTRA domains of TamA, through interaction with POTRA 2, and POTRA-passenger domain interactions mediate passage of the AT passenger domain into the TamA barrel, from where it is released onto the cell surface. The AT barrel domain, which is thought to be inserted into the OM by the BAM complex is omitted for clarity.

8 References

1 Luirink, J., Heijne, G.v., Houben, E., & Gier, J.-W.d. Biogenesis of inner membrane proteins in *Escherichia coli*. *Annual Review of Microbiology* **59**, 329-355 (2005).

2 Fröderberg, L., Houben, E., Samuelson, J. C., Chen, M., Park, S. K., Phillips, G. J., ... & De Gier, J. W. L. Versatility of inner membrane protein biogenesis in *Escherichia coli*. *Molecular Microbiology* **47**, 1015-1027 (2003).

3 Hutchings, M.I., Palmer, T., Harrington, D.J., & Sutcliffe, I.C. Lipoprotein biogenesis in Gram-positive bacteria: knowing when to hold 'em, knowing when to fold 'em. *Trends in Microbiology* **17**, 13-21 (2009).

4 Typas, A., Banzhaf, M., Gross, C.A., & Vollmer, W. From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nature Reviews Microbiology* **10**, 123-136 (2012).

5 Typas, A., Banzhaf, M., Gross, C.A., & Vollmer, W. From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nature Reviews Microbiology* **10**, 123-136 (1998).

6 Nikaido, H. Outer membrane barrier as a mechanism of antimicrobial resistance. *Antimicrobial Agents and Chemotherapy* **33**, 1831 (1989).

7 Koebnik, R., Locher, K.P., & Van Gelder, P. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Molecular Microbiology* **37**, 239-253 (2000).

8 Klemm, P. & Schembri, M.A. Bacterial adhesins: function and structure. *International Journal of Medical Microbiology* **290**, 27-35 (2000).

9 White, D.A., Lennarz, W., & Schnaitman, C.A. Distribution of lipids in the wall and cytoplasmic membrane subfractions of the cell envelope of *Escherichia coli*. *Journal of Bacteriology* **109**, 686-690 (1972).

10 Nikaido, H. Prevention of drug access to bacterial targets: permeability barriers and active efflux. *Science* **264**, 382-388 (1994).

11 Walker, S.L., Redman, J.A., & Elimelech, M. Role of cell surface lipopolysaccharides in *Escherichia coli* K12 adhesion and transport. *Langmuir* **20**, 7736-7746 (2004).

12 Hajjar, A. M., Ernst, R. K., Tsai, J. H., Wilson, C. B., & Miller, S. I. Human Toll-like receptor 4 recognizes host-specific LPS modifications. *Nature Immunology* **3**, 354-359 (2002).

13 Zhou, Z., White, K. A., Polissi, A., Georgopoulos, C., & Raetz, C. R. Function of Escherichia coli MsbA, an essential ABC family transporter, in lipid A and phospholipid biosynthesis. *Journal of Biological Chemistry* **273**, 12466-12475 (1998).

14 Okuda, S., Freinkman, E., & Kahne, D. Cytoplasmic ATP hydrolysis powers transport of lipopolysaccharide across the periplasm in E. coli. *Science* **338**, 1214-1217 (2012).

15 Chng, S.-S., Gronenberg, L.S., & Kahne, D. Proteins required for lipopolysaccharide assembly in *Escherichia coli* form a transenvelope complex. *Biochemistry* **49**, 4565-4567 (2010).

16 Qiao, S., Luo, Q., Zhao, Y., Zhang, X. C., & Huang, Y. Structural basis for lipopolysaccharide insertion in the bacterial outer membrane. *Nature* **511**, 108-111 (2014).

17 Tamm, L.K., Hong, H., & Liang, B. Folding and assembly of β-barrel membrane proteins. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1666**, 250-263 (2004).

18 Van Gelder, P., Saint, N., Van Boxtel, R., Rosenbusch, J. P., & Tommassen, J. Pore functioning of outer membrane protein PhoE of *Escherichia coli*: mutagenesis of the constriction loop L3. *Protein Engineering* **10**, 699-706 (1997).

19 Phale, P. S., Philippsen, A., Widmer, C., Phale, V. P., Rosenbusch, J. P., & Schirmer, T. Role of charged residues at the OmpF porin channel constriction probed by mutagenesis and simulation. *Biochemistry* **40**, 6319-6325 (2001).

20 Nikaido, H. Molecular basis of bacterial outer membrane permeability revisited. *Microbiology and Molecular Biology Reviews* **67**, 593-656 (2003).

21 Cowan, S. W., Schirmer, T., Rummel, G., Steiert, M., Ghosh, R., Pauptit, R. A., ... & Rosenbusch, J. P. Crystal structures explain functional properties of two *E. coli* porins. *Nature* **358**, 727-733 (1992).

22 Bauer, K., Struyve, M., Bosch, D., Benz, R., & Tommassen, J. One single lysine residue is responsible for the special interaction between polyphosphate and the outer membrane porin PhoE of *Escherichia coli*. *Journal of Biological Chemistry* **264**, 16393-16398 (1989).

23 Klebba, P., Hofnung, M. & Charbit, A., A model of maltodextrin transport through the sugar-specific porin, LamB, based on deletion analysis. *The EMBO Journal* **13**, 4670 (1994).

24 Schirmer, T., Keller, T.A., Wang, Y.-F., & Rosenbusch, J.P. Structural basis for sugar translocation through maltoporin channels at 3.1 A resolution. *Science* **267**, 512-514 (1995).

25 Ferguson, A.D. & Deisenhofer, J. TonB-dependent receptors—structural perspectives. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1565**, 318-332 (2002).

26 Neilands, J.B. Siderophores: structure and function of microbial iron transport compounds. *Journal of Biological Chemistry* **270**, 26723-26726 (1995).

27 Noinaj, N., Guillier, M., Barnard, T.J., & Buchanan, S.K. TonB-dependent transporters: regulation, structure, and function. *Annual Review of Microbiology* **64**, 43 (2010).

28 Pawelek, P. D., Croteau, N., Ng-Thow-Hing, C., Khursigara, C. M., Moiseeva, N., Allaire, M., & Coulton, J. W. Structure of TonB in complex with FhuA, *E. coli* outer membrane receptor. *Science* **312**, 1399-1402 (2006).

29 Sklar, J.G., Wu, T., Kahne, D., & Silhavy, T.J. Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in *Escherichia coli*. *Genes & Development* **21**, 2473-2484 (2007).

30 Kim, S., Malinverni, J. C., Sliz, P., Silhavy, T. J., Harrison, S. C., & Kahne, D. Structure and function of an essential component of the outer membrane protein assembly machine. *Science* **317**, 961-964 (2007).

31 Moon, C. P., Zaccai, N. R., Fleming, P. J., Gessmann, D., & Fleming, K. G. Membrane protein thermodynamic stability may serve as the energy sink for sorting in the periplasm. *Proceedings of the National Academy of Sciences* **110**, 4285-4290 (2013).

32 Alba, B.M., Zhong, H.J., Pelayo, J.C., & Gross, C.A. degS (hhoB) is an essential *Escherichia coli* gene whose indispensable function is to provide σE activity. *Molecular Microbiology* **40**, 1323-1333 (2001).

33 Alba, B. M., Leeds, J. A., Onufryk, C., Lu, C. Z., & Gross, C. A. DegS and YaeL participate sequentially in the cleavage of RseA to activate the ςE-dependent extracytoplasmic stress response. *Genes & Development* **16**, 2156-2168 (2002).

34 Ruiz, N. & Silhavy, T.J. Sensing external stress: watchdogs of the *Escherichia coli* cell envelope. *Current Opinion in Microbiology* **8**, 122-126 (2005).

35 Danese, P.N. & Silhavy, T.J. The sigma (E) and the Cpx signal transduction systems control the synthesis of periplasmic protein-folding enzymes in *Escherichia coli*. *Genes & Development* **11**, 1183-1193 (1997).

36 Rouviere, P. E., De Las Penas, A., Mecsas, J., Lu, C. Z., Rudd, K. E., & Gross, C. A. rpoE, the gene encoding the second heat-shock sigma factor, sigma E, in *Escherichia coli*. *The EMBO Journal* **14**, 1032 (1995).

37 Schäfer, U., Beck, K., & Müller, M. Skp, a molecular chaperone of gram-negative bacteria, is required for the formation of soluble periplasmic intermediates of outer membrane proteins. *Journal of Biological Chemistry* **274**, 24567-24574 (1999).

38 Bulieris, P.V., Behrens, S., Holst, O., & Kleinschmidt, J.H. Folding and insertion of the outer membrane protein OmpA is assisted by the chaperone Skp and by lipopolysaccharide. *Journal of Biological Chemistry* **278**, 9092-9099 (2003).

39 Walton, T.A. & Sousa, M.C. Crystal structure of Skp, a prefoldin-like chaperone that protects soluble and membrane proteins from aggregation. *Molecular Cell* **15**, 367-374 (2004).

40 Burmann, B.M., Wang, C., & Hiller, S. Conformation and dynamics of the periplasmic membrane-protein–chaperone complexes OmpX–Skp and tOmpA–Skp. *Nature Structural Molecular Biology* **20**, 1265-1272 (2013).

41 Chen, R. & Henning, U. Aperiplasmic protein (Skp) of *Escherichia coli* selectively binds a class of outer membrane proteins. *Molecular Microbiology* **19**, 1287-1294 (1996).

42 Bitto, E. & McKay, D.B. Crystallographic structure of SurA, a molecular chaperone that facilitates folding of outer membrane porins. *Structure* **10**, 1489-1498 (2002).

43 Justice, S. S., Hunstad, D. A., Harper, J. R., Duguay, A. R., Pinkner, J. S., Bann, J., ... & Hultgren, S. J. Periplasmic peptidyl prolyl cis-trans isomerases are not essential for viability, but SurA is required for pilus biogenesis in *Escherichia coli*. *Journal of Bacteriology* **187**, 7680-7686 (2005).

44 Bitto, E. & McKay, D.B. The periplasmic molecular chaperone protein SurA binds a peptide motif that is characteristic of integral outer membrane proteins. *Journal of Biological Chemistry* **278**, 49316-49322 (2003).

45 Xu, X., Wang, S., Hu, Y.-X., & McKay, D.B. The periplasmic bacterial molecular chaperone SurA adapts its structure to bind peptides in different conformations to assert a sequence preference for aromatic residues. *Journal of Molecular Biology* **373**, 367-381 (2007).

46 Krojer, T., Garrido-Franco, M., Huber, R., Ehrmann, M., & Clausen, T. Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine. *Nature* **416**, 455-459 (2002).

47 Krojer, T., Sawa, J., Schäfer, E., Saibil, H. R., Ehrmann, M., & Clausen, T. Structural basis for the regulated protease and chaperone function of DegP. *Nature* **453**, 885-890 (2008).

48 Genevrois, S., Steeghs, L., Roholl, P., Letesson, J. J., & van der Ley, P. The Omp85 protein of *Neisseria meningitidis* is required for lipid export to the outer membrane. *The EMBO Journal* **22**, 1780-1789 (2003).

49 Voulhoux, R., Bos, M. P., Geurtsen, J., Mols, M., & Tommassen, J. Role of a highly conserved bacterial protein in outer membrane protein assembly. *Science* **299**, 262-265 (2003).

50 Wu, T., Malinverni, J., Ruiz, N., Kim, S., Silhavy, T. J., & Kahne, D. Identification of a multicomponent complex required for outer membrane biogenesis in *Escherichia coli*. *Cell* **121**, 235-245 (2005).

51 Onufryk, C., Crouch, M.-L., Fang, F.C., & Gross, C.A. Characterization of six lipoproteins in the σE regulon. *Journal of Bacteriology* **187**, 4552-4561 (2005).

52 Charlson, E.S., Werner, J.N., & Misra, R. Differential effects of yfgL mutation on *Escherichia coli* outer membrane proteins and lipopolysaccharide. *Journal of Bacteriology* **188**, 7186-7194 (2006).

53 Malinverni, J. C., Werner, J., Kim, S., Sklar, J. G., Kahne, D., Misra, R., & Silhavy, T. J. YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in *Escherichia coli*. *Molecular Microbiology* **61**, 151-164 (2006).

54 Hagan, C.L., Silhavy, T.J., & Kahne, D. β-barrel membrane protein assembly by the Bam complex. *Annual Review of Biochemistry* **80**, 189-210 (2011).

55 Hagan, C.L., Kim, S., & Kahne, D. Reconstitution of outer membrane protein assembly from purified components. *Science* **328**, 890-892 (2010).

56 Robert, V., Volokhina, E. B., Senf, F., Bos, M. P., Van Gelder, P., & Tommassen, J Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif. *PLoS Biol* **4**, e377 (2006).

57 Bodelón, G., Marín, E., & Fernández, L.Á. Role of periplasmic chaperones and BamA (YaeT/Omp85) in folding and secretion of intimin from enteropathogenic *Escherichia coli* strains. *Journal of Bacteriology* **191**, 5169-5179 (2009).

58 Knowles, T. J., Jeeves, M., Bobat, S., Dancea, F., McClelland, D., Palmer, T., ... & Henderson, I. R. Fold and function of polypeptide transport-associated domains responsible for delivering unfolded proteins to membranes. *Molecular Microbiology* **68**, 1216-1227 (2008).

59 Bennion, D., Charlson, E.S., Coon, E., & Misra, R. Dissection of β-barrel outer membrane protein assembly pathways through characterizing BamA POTRA 1 mutants of *Escherichia coli*. *Molecular Microbiology* **77**, 1153-1171 (2010).

60 Ieva, R. & Bernstein, H.D. Interaction of an autotransporter passenger domain with BamA during its translocation across the bacterial outer membrane. *Proceedings of the National Academy of Sciences* **106**, 19120-19125 (2009).

61 Noinaj, N., Kuszak, A. J., Gumbart, J. C., Lukacik, P., Chang, H., Easley, N. C., ... & Buchanan, S. K. Structural insight into the biogenesis of beta-barrel membrane proteins. *Nature* **501**, 385-390 (2013).

62 Albrecht, R., Schutz, M., Oberhettinger, P., Faulstich, M., Bermejo, I., Rudel, T., ... & Zeth, K. Structure of BamA, an essential factor in outer membrane protein biogenesis. *Acta Crystallographica Section D: Biological Crystallography* **70**, 1779-1789 (2014).

63 Kleinschmidt, J.H. & Tamm, L.K. Secondary and tertiary structure formation of the β-barrel membrane protein OmpA is synchronized and depends on membrane thickness. *Journal of Molecular Biology* **324**, 319-330 (2002).

64 Hagan, C.L., Westwood, D.B., & Kahne, D. Bam lipoproteins assemble BamA in vitro. *Biochemistry* **52**, 6108-6113 (2013).

65 Misra, R., Stikeleather, R., & Gabriele, R. *In vivo* roles of BamA, BamB and BamD in the biogenesis of BamA, a core protein of the β-barrel assembly machine of *Escherichia coli*. *Journal of Molecular Biology* (2014).

66 Noinaj, N., Fairman, J.W., & Buchanan, S.K. The Crystal structure of BamB suggests interactions with BamA and its role within the BAM complex. *Journal of Molecular Biology* **407**, 248-260 (2011).

67 Kim, K.H. & Paetzel, M. Crystal structure of *Escherichia coli* BamB, a lipoprotein component of the β-barrel assembly machinery complex. *Journal of Molecular Biology* **406**, 667-678 (2011).

68 Smith, T.F. Diversity of WD-repeat proteins in *The coronin family of proteins* (Springer, 2008), pp. 20-30.

69 Vuong, P., Bennion, D., Mantei, J., Frost, D., & Misra, R. Analysis of YfgL and YaeT interactions through bioinformatics, mutagenesis, and biochemistry. *Journal of Bacteriology* **190**, 1507-1517 (2008).

70 Heuck, A., Schleiffer, A., & Clausen, T. Augmenting β-augmentation: structural basis of how BamB binds BamA and may support folding of outer membrane proteins. *Journal of Molecular Biology* **406**, 659-666 (2011).

71 Warner, L. R., Varga, K., Lange, O. F., Baker, S. L., Baker, D., Sousa, M. C., & Pardi, A. Structure of the BamC two-domain protein obtained by Rosetta with a limited NMR data set. *Journal of Molecular Biology* **411**, 83-95 (2011).

72 Albrecht, R. & Zeth, K. Structural basis of outer membrane protein biogenesis in bacteria. *Journal of Biological Chemistry* **286**, 27792-27803 (2011).

73 Kim, K.H., Aulakh, S., & Paetzel, M. Crystal Structure of β-Barrel Assembly Machinery BamCD Protein Complex. *Journal of Biological Chemistry* **286**, 39116-39121 (2011).

74 Webb, C. T., Selkrig, J., Perry, A. J., Noinaj, N., Buchanan, S. K., & Lithgow, T. Dynamic association of BAM complex modules includes surface exposure of the lipoprotein BamC. *Journal of Molecular Biology* **422**, 545-555 (2012).

75 Gatsos, X., Perry, A. J., Anwari, K., Dolezal, P., Wolynec, P. P., Likić, V. A., ... & Lithgow, T. Protein secretion and outer membrane assembly in alphaproteobacteria. *FEMS microbiology reviews* **32**, 995-1009 (2008).

76 Sklar, J. G., Wu, T., Gronenberg, L. S., Malinverni, J. C., Kahne, D., & Silhavy, T. J. Lipoprotein SmpA is a component of the YaeT complex that assembles outer membrane proteins in *Escherichia coli*. *Proceedings of the National Academy of Sciences* **104**, 6400-6405 (2007).

77 Knowles, T. J., Browning, D. F., Jeeves, M., Maderbocus, R., Rajesh, S., Sridhar, P., ... & Overduin, M. Structure and function of BamE within the outer membrane and the β-barrel assembly machine. *EMBO reports* **12**, 123-128 (2011).

78 Kim, K. H., Kang, H. S., Okon, M., Escobar-Cabrera, E., McIntosh, L. P., & Paetzel, M. Structural characterization of Escherichia coli BamE, a lipoprotein component of the β-barrel assembly machinery complex. *Biochemistry* **50**, 1081-1090 (2011).

79 Reumann, S., Davila-Aponte, J., & Keegstra, K. The evolutionary origin of the protein-translocating channel of chloroplastic envelope membranes: Identification of a cyanobacterial homolog. *Proceedings of the National Academy of Sciences* **96**, 784-789 (1999).

80 Hsu, S. C., Patel, R., Bédard, J., Jarvis, P., & Inoue, K. Two distinct Omp85 paralogs in the chloroplast outer envelope membrane are essential for embryogenesis in *Arabidopsis thaliana*. *Plant Signal Behaviour* **3**, 1134-1135 (2008).

81 Künkele, K. P., Heins, S., Dembowski, M., Nargang, F. E., Benz, R., Thieffry, M., ... & Neupert, W. The preprotein translocation channel of the outer membrane of mitochondria. *Cell* **93**, 1009-1019 (1998).

82 Wiedemann, N., Kozjak, V., Chacinska, A., Schönfisch, B., Rospert, S., Ryan, M. T., ... & Meisinger, C. Machinery for protein sorting and assembly in the mitochondrial outer membrane. *Nature* **424**, 565-571 (2003).

83 Ishikawa, D., Yamamoto, H., Tamura, Y., Moritoh, K., & Endo, T. Two novel proteins in the mitochondrial outer membrane mediate β-barrel protein assembly. *The Journal of cell biology* **166**, 621-627 (2004).

84 Waizenegger, T., Habib, S. J., Lech, M., Mokranjac, D., Paschen, S. A., Hell, K., ... & Rapaport, D. Tob38, a novel essential component in the biogenesis of β-barrel proteins of mitochondria. *EMBO reports* **5**, 704-709 (2004).

85 Kozjak, V., Wiedemann, N., Milenkovic, D., Lohaus, C., Meyer, H. E., Guiard, B., ... & Pfanner, N. An essential role of Sam50 in the protein sorting and assembly machinery of the mitochondrial outer membrane. *Journal of Biological Chemistry* **278**, 48520-48523 (2003).

86 Milenkovic, D., Kozjak, V., Wiedemann, N., Lohaus, C., Meyer, H. E., Guiard, B., ... & Meisinger, C. Sam35 of the mitochondrial protein sorting and assembly machinery is a peripheral outer membrane protein essential for cell viability. *Journal of Biological Chemistry* **279**, 22781-22785 (2004).

87 Kutik, S., Stojanovski, D., Becker, L., Becker, T., Meinecke, M., Krüger, V., ... & Wiedemann, N. Dissecting membrane insertion of mitochondrial β-barrel proteins. *Cell* **132**, 1011-1024 (2008).

88 Sánchez-Pulido, L., Devos, D., Genevrois, S., Vicente, M., & Valencia, A. POTRA: a conserved domain in the FtsQ family and a class of β-barrel outer membrane proteins. *Trends in Biochemical Sciences* **28**, 523-526 (2003).

89 Van Den Ent, F., Vinkenvleugel, T. M., Ind, A., West, P., Veprintsev, D., Nanninga, N., ... & Löwe, J. Structural and mutational analysis of the cell division protein FtsQ. *Molecular Microbiology* **68**, 110-123 (2008).

90 Arnold, T., Zeth, K., & Linke, D. Omp85 from the Thermophilic cyanobacterium thermosynechococcus elongatus differs from proteobacterial Omp85 in structure and domain composition. *Journal of Biological Chemistry* **285**, 18003-18015 (2010).

91 Milner, D. S., Till, R., Cadby, I., Lovering, A. L., Basford, S. M., Saxon, E. B., ... & Sockett, R. E. Ras GTPase-like protein MglA, a controller of bacterial social-motility in Myxobacteria, has evolved to control bacterial predation by Bdellovibrio. *PLoS genetics* **10**, e1004253 (2014).

92 Bos, M.P., Robert, V., & Tommassen, J. Functioning of outer membrane protein assembly factor Omp85 requires a single POTRA domain. *EMBO reports* **8**, 1149-1154 (2007).

93 Koenig, P., Mirus, O., Haarmann, R., Sommer, M. S., Sinning, I., Schleiff, E., & Tews, I. Conserved properties of polypeptide transport-associated (POTRA) domains derived from Cyanobacterial Omp85. *Journal of Biological Chemistry* **285**, 18016-18024 (2010).

94 Delattre, A. S., Saint, N., Clantin, B., Willery, E., Lippens, G., Locht, C., ... & Jacob-Dubuisson, F. Substrate recognition by the POTRA domains of TpsB transporter FhaC. *Molecular Microbiology* **81**, 99-112 (2011).

95 Gatzeva-Topalova, P.Z., Warner, L.R., Pardi, A., & Sousa, M.C. Structure and flexibility of the complete periplasmic domain of BamA: the protein insertion machine of the outer membrane. *Structure* **18**, 1492-1501 (2008).

96 Misra, R., First glimpse of the crystal structure of YaeT's POTRA domains. *ACS Chemical Biology* **2**, 649-651 (2007).

97 Hodak, H., Clantin, B., Willery, E., Villeret, V., Locht, C., & Jacob-Dubuisson, F. Secretion signal of the filamentous haemagglutinin, a model two-partner secretion substrate. *Molecular Microbiology* **61**, 368-382 (2006).

98 Bos, M.P., Robert, V., & Tommassen, J. Functioning of outer membrane protein assembly factor Omp85 requires a single POTRA domain. *EMBO reports* **8**, 1149-1154 (2007).

99 Habib, S. J., Waizenegger, T., Niewienda, A., Paschen, S. A., Neupert, W., & Rapaport, D. The N-terminal domain of Tob55 has a receptor-like function in the biogenesis of mitochondrial β-barrel proteins. *The Journal of cell biology* **176**, 77-88 (2007).

100 Stroud, D. A., Becker, T., Qiu, J., Stojanovski, D., Pfannschmidt, S., Wirth, C., ... & Wiedemann, N. Biogenesis of mitochondrial β-barrel proteins: the POTRA domain is involved in precursor release from the SAM complex. *Molecular Biology of the Cell* **22**, 2823-2833 (2011).

101 Henderson, I. R., Navarro-Garcia, F., Desvaux, M., Fernandez, R. C., & Ala'Aldeen, D. Type V protein secretion pathway: the autotransporter story. *Microbiology and Molecular Biology Reviews* **68**, 692-744 (2004).

102 Cotter, P. A., Yuk, M. H., Mattoo, S., Akerley, B. J., Boschwitz, J., Relman, D. A., & Miller, J. F. Filamentous hemagglutinin of Bordetella bronchiseptica Is required for efficient establishment of tracheal colonization. *Infection and Immunity* **66**, 5921-5929 (1998).

103 Coutte, L., Antoine, R., Drobecq, H., Locht, C., & Jacob-Dubuisson, F. Subtilisin-like autotransporter serves as maturation protease in a bacterial secretion pathway. *The EMBO Journal* **20**, 5040-5048 (2001).

104 Clantin, B., Hodak, H., Willery, E., Locht, C., Jacob-Dubuisson, F., & Villeret, V. The crystal structure of filamentous hemagglutinin secretion domain and its implications for the two-partner secretion pathway. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 6194-6199 (2004).

105 Clantin, B., Delattre, A. S., Rucktooa, P., Saint, N., Méli, A. C., Locht, C., ... & Villeret, V. Structure of the membrane protein FhaC: a member of the Omp85-TpsB transporter superfamily. *Science* **317**, 957-961 (2007).

106 Guédin, S., Willery, E., Tommassen, J., Fort, E., Drobecq, H., Locht, C., & Jacob-Dubuisson, F. Novel topological features of FhaC, the outer membrane transporter involved in the secretion of the *Bordetella pertussis* filamentous hemagglutinin. *Journal of Biological Chemistry* **275**, 30202-30210 (2000).

107 Loveless, B.J. & Saier, M.H. A novel family of channel-forming, autotransporting, bacterial virulence factors. *Molecular membrane biology* **14**, 113-123 (1997).

108 Sijbrandi, R., Urbanus, M. L., Corinne, M., Bernstein, H. D., Oudega, B., Otto, B. R., & Luirink, J. Signal recognition particle (SRP)-mediated targeting and Sec-dependent translocation of an extracellular *Escherichia coli* protein. *Journal of Biological Chemistry* **278**, 4654-4659 (2003).

109 Szabady, R.L., Peterson, J.H., Skillman, K.M., & Bernstein, H.D. An unusual signal peptide facilitates late steps in the biogenesis of a bacterial autotransporter. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 221-226 (2005).

110 Peterson, J.H., Szabady, R.L., & Bernstein, H.D. An unusual signal peptide extension inhibits the binding of bacterial presecretory proteins to the signal recognition particle, trigger factor, and the SecYEG complex. *Journal of Biological Chemistry* **281**, 9038-9048 (2006).

111 Jong, W. S., Corinne, M., Ruijter, E., Orru, R. V., Genevaux, P., & Luirink, J. YidC is involved in the biogenesis of the secreted autotransporter hemoglobin protease. *Journal of Biological Chemistry* **285**, 39682-39690 (2010).

112 Purdy, G.E., Fisher, C.R., & Payne, S.M., IcsA surface presentation in *Shigella flexneri* requires the periplasmic chaperones DegP, Skp, and SurA. *Journal of Bacteriology* **189**, 5566-5573 (2007).

113 Wagner, J. K., Heindl, J. E., Gray, A. N., Jain, S., & Goldberg, M. B Contribution of the periplasmic chaperone Skp to efficient presentation of the autotransporter IcsA on the surface of *Shigella flexneri*. *Journal of Bacteriology* **191**, 815-821 (2009).

114 Ieva, R., Tian, P., Peterson, J.H., & Bernstein, H.D. Sequential and spatially restricted interactions of assembly factors with an autotransporter β domain. *Proceedings of the National Academy of Sciences* **108**, E383-E391 (2011).

115 Ruiz-Perez, F., Henderson, I. R., Leyton, D. L., Rossiter, A. E., Zhang, Y., & Nataro, J. P. Roles of periplasmic chaperone proteins in the biogenesis of serine protease autotransporters of Enterobacteriaceae. *Journal of Bacteriology* **191**, 6571-6583 (2009).

116 Sauri, A., Soprova, Z., Wickström, D., de Gier, J. W., Van der Schors, R. C., Smit, A. B., ... & Luirink, J. The Bam (Omp85) complex is involved in secretion of the autotransporter haemoglobin protease. *Microbiology* **155**, 3982-3991 (2009).

117 Oomen, C. J., van Ulsen, P., Van Gelder, P., Feijen, M., Tommassen, J., & Gros, P. Structure of the translocator domain of a bacterial autotransporter. *The EMBO Journal* **23**, 1257-1266 (2004).

118 Khalid, S. & Sansom, M.S. Molecular dynamics simulations of a bacterial autotransporter: NalP from Neisseria meningitidis. *Molecular membrane biology* **23**, 499-508 (2006).

119 Barnard, T. J., Dautin, N., Lukacik, P., Bernstein, H. D., & Buchanan, S. K. Autotransporter structure reveals intra-barrel cleavage followed by conformational changes. *Nat Struct Mol Biol* **14**, 1214-1220 (2007).

120 Zhai, Y., Zhang, K., Huo, Y., Zhu, Y., Zhou, Q., Lu, J., ... & Sun, F. Autotransporter passenger domain secretion requires a hydrophobic cavity at the extracellular entrance of the β-domain pore. *Biochemical Journal* **435**, 577-587 (2011).

121 Veiga, E., Sugawara, E., Nikaido, H., de Lorenzo, V., & Fernández, L. A. Export of autotransported proteins proceeds through an oligomeric ring shaped by C-terminal domains. *The EMBO Journal* **21**, 2122-2131 (2002).

122 Van Ulsen, P., Van Alphen, L., Ten Hove, J., Fransen, F., Van Der Ley, P., & Tommassen, J. A Neisserial autotransporter NalP modulating the processing of other autotransporters. *Molecular Microbiology* **50**, 1017-1030 (2003).

123 Dautin, N., Barnard, T.J., Anderson, D.E., & Bernstein, H.D. Cleavage of a bacterial autotransporter by an evolutionarily convergent autocatalytic mechanism. *The EMBO Journal* **26**, 1942-1952 (2007).

124 Barnard, T. J., Gumbart, J., Peterson, J. H., Noinaj, N., Easley, N. C., Dautin, N., ... & Buchanan, S. K. Molecular basis for the activation of a catalytic asparagine residue in a self-cleaving bacterial autotransporter. *Journal of Molecular Biology* **415**, 128-142 (2012).

125 Celik, N., Webb, C. T., Leyton, D. L., Holt, K. E., Heinz, E., Gorrell, R., ... & Lithgow, T. A bioinformatic strategy for the detection, classification and analysis of bacterial autotransporters. *PLoS ONE* **7**, e43245 (2012).

126 Leyton, D. L., Johnson, M. D., Thapa, R., Huysmans, G. H., Dunstan, R. A., Celik, N., ... & Lithgow, T. A mortise–tenon joint in the transmembrane domain modulates autotransporter assembly into bacterial outer membranes. *Nat Commun* **5** (2014).

127 Junker, M., Schuster, C. C., McDonnell, A. V., Sorg, K. A., Finn, M. C., Berger, B., & Clark, P. L. Pertactin β-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 4918-4923 (2006).

128 Nishimura, K., Tajima, N., Yoon, Y. H., Park, S. Y., & Tame, J. R. Autotransporter passenger proteins: virulence factors with common structural themes. *Journal of Molecular Medicine* **88**, 451-458 (2010).

129 van den Berg, B. Crystal structure of a full-length autotransporter. *Journal of Molecular Biology* **396**, 627-633 (2011).

130 Heras, B., Totsika, M., Peters, K. M., Paxman, J. J., Gee, C. L., Jarrott, R. J., ... & Schembri, M. A. The antigen 43 structure reveals a molecular Velcro-like mechanism of autotransporter-mediated bacterial clumping. *Proceedings of the National Academy of Sciences,* **111,** 457-462 (2013).

131 Junker, M., Besingi, R.N., & Clark, P.L. Vectorial transport and folding of an autotransporter virulence protein during outer membrane secretion. *Molecular Microbiology* **71**, 1323-1332 (2009).

132 Junker, M. & Clark, P.L. Slow formation of aggregation-resistant β-sheet folding intermediates. *Proteins: Structure, Function, and Bioinformatics* **78**, 812-824 (2010).

133 Renn, J. P., Junker, M., Besingi, R. N., Braselmann, E., & Clark, P. L. ATP-Independent control of autotransporter virulence protein transport via the folding properties of the secreted protein. *Chemistry & Biology* (2012).

134 Yeo, H. J., Cotter, S. E., Laarmann, S., Juehne, T., St Geme, J. W., & Waksman, G. Structural basis for host recognition by the *Haemophilus influenzae* Hia autotransporter. *The EMBO Journal* **23**, 1245-1256 (2004).

135 Shahid, S. A., Bardiaux, B., Franks, W. T., Krabben, L., Habeck, M., van Rossum, B. J., & Linke, D. Membrane-protein structure determination by solid-state NMR spectroscopy of microcrystals. *Nature Methods* **9**, 1212-1217 (2012).

136 Jong, W. S., Hagen-Jongman, T., Corinne, M., Den Blaauwen, T., Jan Slotboom, D., Tame, J. R., ... & Luirink, J. Limited tolerance towards folded elements during secretion of the autotransporter Hbp. *Molecular Microbiology* **63**, 1524-1536 (2007).

137 Veiga, E., Sugawara, E., Nikaido, H., de Lorenzo, V., & Fernández, L. A. Export of autotransported proteins proceeds through an oligomeric ring shaped by C-terminal domains. *The EMBO journal*, ***21***, 2122-2131 (2002).

138 Selkrig, J., Mosbahi, K., Webb, C. T., Belousoff, M. J., Perry, A. J., Wells, T. J., ... & Lithgow, T. Discovery of an archetypal protein transport system in bacterial outer membranes. *Nature Structural and Molecular Biology* **19**, 506-510 (2012).

139 Burall, L. S., Harro, J. M., Li, X., Lockatell, C. V., Himpsl, S. D., Hebel, J. R., ... & Mobley, H. L. Proteus mirabilis genes that contribute to pathogenesis of urinary tract infection: identification of 25 signature-tagged mutants attenuated at least 100-Fold. *Infection and Immunity* **72**, 2922-2938 (2004).

140 Gruss, F., Zähringer, F., Jakob, R. P., Burmann, B. M., Hiller, S., & Maier, T. The structural basis of autotransporter translocation by TamA. *Nature Structural and Molecular Biology* **20**, 1318-1320 (2013).

141 Noinaj, N., Kuszak, A. J., Balusek, C., Gumbart, J. C., & Buchanan, S. K. Lateral opening and exit pore formation are required for BamA function. *Structure* **22**, 1055-1062 (2014).

142 Leonard-Rivera, M. & Misra, R., Conserved residues of the putative L6 loop of *Escherichia coli* BamA play a critical role in the assembly of β-barrel outer membrane proteins, including that of BamA itself. *Journal of Bacteriology* **194**, 4662-4668 (2012).

143 Pusnik, M., Schmidt, O., Perry, A. J., Oeljeklaus, S., Niemann, M., Warscheid, B., ... & Schneider, A. Mitochondrial preprotein translocase of trypanosomatids has a bacterial origin. *Current Biology* **21**, 1738-1743 (2011).

144 Pusnik, M., Mani, J., Schmidt, O., Niemann, M., Oeljeklaus, S., Schnarwiler, F., ... & Schneider, A. An essential novel component of the noncanonical mitochondrial outer membrane protein import system of trypanosomatids. *Molecular Biology of the Cell* **23**, 3420-3428 (2012).

145 Zarsky, V., Tachezy, J., & Dolezal, P. Tom40 is likely common to all mitochondria. *Current Biology* **22**, R479-R481 (2012).

146 Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., ... & Mori, H. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology* **2** (2006).

147 Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., ... & Laue, E. D. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins: Structure, Function, and Bioinformatics* **59**, 687-696 (2005).

148 Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., & Clore, G. M. Overcoming the overlap problem in the assignment of proton NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1. beta. *Biochemistry* **28**, 6150-6156 (1989).

149 Whitmore, L. & Wallace, B.A. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* **89**, 392-400 (2008).
150 Provencher, S.W. & Gloeckner, J. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **20**, 33-37 (1981).
151 Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophysical Journal* **78**, 1606-1619 (2000).
152 Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., ... & Svergun, D. I. New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of Applied Crystallography* **45**, 342-350 (2012).
153 Volkov, V.V. & Svergun, D.I. Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography* **36**, 860-864 (2003).
154 Svergun, D., Barberato, C., & Koch, M.H.J. CRYSOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography* **28**, 768-773 (1995).
155 Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., & Svergun, D. I. Structural characterization of flexible proteins using small-angle X-ray scattering. *Journal of the American Chemical Society* **129**, 5656-5664 (2007).
156 Brookes, E., Demeler, B., Rosano, C., & Rocco, M. The implementation of SOMO (SOlution MOdeller) in the UltraScan analytical ultracentrifugation data analysis suite: enhanced capabilities allow the reliable hydrodynamic modeling of virtually any kind of biomacromolecule. *European Biophysics Journal* **39**, 423-435 (2010).
157 Petoukhov, M.V. & Svergun, D.I. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophysical Journal* **89**, 1237-1250 (2005).
158 Hammel, M. Validation of macromolecular flexibility in solution by small-angle X-ray scattering (SAXS). *European Biophysics Journal* **41**, 789-799 (2010).
159 Rambo, R.P. & Tainer, J.A. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95**, 559-571 (2008).
160 Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., ... & Wilson, K. S. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* **67**, 235-242 (2011).
161 Battye, T.G.G. *et al.*, iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D: Biological Crystallography* **67**, 271-281 (2011).
162 Evans, P.R. & Murshudov, G.N. How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography* **69**, 0-0 (2013).
163 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. Phaser crystallographic software. *Journal of Applied Crystallography* **40**, 658-674 (2007).
164 Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., ... & Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography* **67**, 355-367 (2011).

165 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* **60**, 2126-2132 (2004).
166 Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Echols, N., Headd, J. J., ... & Zwart, P. H. The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94-106 (2011).
167 Ward, R., Zoltner, M., Beer, L., El Mkami, H., Henderson, I. R., Palmer, T., & Norman, D. G. The orientation of a tandem POTRA domain pair, of the beta-barrel assembly protein BamA, determined by PELDOR spectroscopy. *Structure* **17**, 1187-1194 (2009).
168 Gabel, F., Lensink, M. F., Clantin, B., Jacob-Dubuisson, F., Villeret, V., & Ebel, C Probing the conformation of FhaC with small-angle neutron scattering and molecular modelling. *Biophysical Journal* **107**, 185-196 (2014).
169 Patel, G.J. & Kleinschmidt, J.H. The lipid bilayer-inserted membrane protein BamA of Escherichia coli facilitates insertion and folding of outer membrane protein A from its complex with Skp. *Biochemistry* **52**, 3974-3986 (2013).
170 Babu, M., Díaz-Mejía, J. J., Vlasblom, J., Gagarinova, A., Phanse, S., Graham, C., ... & Emili, A. Genetic interaction maps in Escherichia coli reveal functional crosstalk among cell envelope biogenesis pathways. *PLoS genetics* **7**, e1002377 (2011).
171 Seufferheld, M.J., Alvarez, H.M., & Farias, M.E. Role of polyphosphates in microbial adaptation to extreme environments. *Applied and Environmental Microbiology* **74**, 5867-5874 (2008).
172 Gallant, C. V., Sedic, M., Chicoine, E. A., Ruiz, T., & Mintz, K. P. Membrane morphology and leukotoxin secretion are associated with a novel membrane protein of *Aggregatibacter actinomycetemcomitans*. *Journal of Bacteriology* **190**, 5972-5980 (2008).
173 Paradies, H.H. Shape and size of a nonionic surfactant micelle. Triton X-100 in aqueous solution. *The Journal of Physical Chemistry* **84**, 599-607 (1980).
174 Slotboom, D.J., Duurkens, R.H., Olieman, K., & Erkens, G.B., Static light scattering to characterize membrane proteins in detergent solution. *Methods* **46**, 73-82 (2008).
175 Dorwart, M. R., Wray, R., Brautigam, C. A., Jiang, Y., & Blount, P. *S. aureus* MscL is a pentamer in vivo but of variable stoichiometries in vitro: implications for detergent-solubilized membrane proteins. *PLoS biology* **8**, e1000555 (2010).
176 Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M., & Barton, G. J. JPred: a consensus secondary structure prediction server. *Bioinformatics* **14**, 892-893 (1998).
177 Xue, Z., Xu, D., Wang, Y., & Zhang, Y. ThreaDom: extracting protein domain boundary information from multiple threading alignments. *Bioinformatics* **29**, i247-i256 (2013).
178 Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., & Russell, R. B Protein disorder prediction: implications for structural proteomics. *Structure* **11**, 1453-1459 (2003).
179 Ishida, T. & Kinoshita, K. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic acids research* **35**, W460-W464 (2007).
180 Hofman, K. TMbase-A database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler* **374**, 166 (1993).

181 Kelley, L.A. & Sternberg, M.J. Protein structure prediction on the Web: a case study using the Phyre server. *Nature protocols* **4**, 363-371 (2009).
182 Roy, A., Kucukural, A., & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols* **5**, 725-738 (2010).
183 Fitzgerald, P. & Madsen, N.B. Improvement of limit of diffraction and useful X-ray lifetime of crystals of glycogen debranching enzyme. *Journal of Crystal Growth* **76**, 600-606 (1986).
184 Dong, A., Xu, X., Edwards, A. M., Chang, C., Chruszcz, M., Cuff, M., ... & Nyman, T. In situ proteolysis for protein crystallization and structure determination. *Nature Methods* **4**, 1019-1021 (2007).
185 Zwart, P., Grosse-Kunstleve, R., & Adams, P. Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation. *CCP4 newsletter* **43**, 27-35 (2005).
186 Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., ... & Schwede, T. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic acids research*, gku340 (2014).
187 Chen, C. P., Posy, S., Ben-Shaul, A., Shapiro, L., & Honig, B. H. Specificity of cell–cell adhesion by classical cadherins: critical role for low-affinity dimerization through β-strand swapping. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 8531-8536 (2005).
188 Kreisberg, J.F., Betts, S.D., & King, J. β-Helix core packing within the triple-stranded oligomerization domain of the P22 tailspike. *Protein Science* **9**, 2338-2343 (2000).
189 Von Heijne, G. Membrane protein structure prediction: hydrophobicity analysis and the positive-inside rule. *Journal of Molecular Biology* **225**, 487-494 (1992).
190 Ricci, D.P., Hagan, C.L., Kahne, D., & Silhavy, T.J. Activation of the *Escherichia coli* β-barrel assembly machine (Bam) is required for essential components to interact properly with substrate. *Proceedings of the National Academy of Sciences* **109**, 3487-3491 (2012).
191 Walsh, N. P., Alba, B. M., Bose, B., Gross, C. A., & Sauer, R. T. OMP peptide signals initiate the envelope-stress response by activating DegS protease via relief of inhibition mediated by its PDZ domain. *Cell* **113**, 61-71 (2003).
192 Pletneva, E.V., Gray, H.B., & Winkler, J.R. Nature of the cytochrome C molten globule. *Journal of the American Chemical Society* **127**, 15370-15371 (2005).
193 Vamvaca, K., Vögeli, B., Kast, P., Pervushin, K., & Hilvert, D. An enzymatic molten globule: efficient coupling of folding and catalysis. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 12860-12864 (2004).
194 Ubbink, M. The courtship of proteins: understanding the encounter complex. *FEBS letters* **583**, 1060-1066 (2009).
195 Van der Woude, M.W. & Henderson, I.R. Regulation and function of Ag43 (flu). *Annual Review of Microbiology* **62**, 153-169 (2008).
196 Behlke, J. & Ristau, O. Molecular mass determination by sedimentation velocity experiments and direct fitting of the concentration profiles. *Biophysical Journal* **72**, 428-434 (1997).
197 Kjaergaard, M., Brander, S., & Poulsen, F.M. Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. *Journal of biomolecular NMR* **49**, 139-149 (2011).

198 Kurzbach, D., Platzer, G., Schwarz, T. C., Henen, M. A., Konrat, R., & Hinderberger, D. Cooperative unfolding of compact conformations of the intrinsically disordered protein osteopontin. *Biochemistry* **52**, 5167-5175 (2013).

199 Boze, H., Marlin, T., Durand, D., Pérez, J., Vernhet, A., Canon, F., ... & Cabane, B. Proline-rich salivary proteins have extended conformations. *Biophysical Journal* **99**, 656-665 (2010).

200 Choi, Ucheor B., McCann, James J., Weninger, Keith R., & Bowen, Mark E. Beyond the random coil: stochastic conformational switching in intrinsically disordered proteins. *Structure* **19**, 566-576 (2011).

201 Bernadó, P. & Svergun, D.I. Analysis of intrinsically disordered proteins by small-angle X-ray scattering in *Intrinsically Disordered Protein Analysis* (Springer, 2012), pp. 107-122.

202 Roman-Hernandez, G., Peterson, J.H., & Bernstein, H.D. Reconstitution of bacterial autotransporter assembly using purified components. *eLife* **3**, e04234 (2014).

203 Henderson, I.R. & Owen, P. The major phase-variable outer membrane protein of Escherichia coli structurally resembles the immunoglobulin A1 protease class of exported protein and is regulated by a novel mechanism involving Dam and OxyR. *Journal of Bacteriology* **181**, 2132-2141 (1999).

204 Volkov, A., Ubbink, M., & van Nuland, N.J. Mapping the encounter state of a transient protein complex by PRE NMR spectroscopy. *Journal of biomolecular NMR* **48**, 225-236 (2010).

205 Iwahara, J. & Clore, G.M. Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature* **440**, 1227-1230 (2006).

206 Panchal, S.C., Bhavesh, N.S., & Hosur, R.V. Improved 3D triple resonance experiments, HNN and HN (C) N, for HN and 15N sequential correlations in (13C, 15N) labeled proteins: application to unfolded proteins. *Journal of biomolecular NMR* **20**, 135-147 (2001).

207 Ruiz-Perez, F., Henderson, I.R., & Nataro, J.P. Interaction of FkpA, a peptidyl-prolyl cis-trans isomerase with EspP autotransporter protein. *Gut Microbes* **1**, 339-344 (2010).

208 Heinz, E. & Lithgow, T. A comprehensive analysis of the Omp85/TpsB protein superfamily structural diversity, taxonomic occurrence, and evolution. *Frontiers in microbiology* **5** (2014).

209 Kang'ethe, W. & Bernstein, H.D. Charge-dependent secretion of an intrinsically disordered protein via the autotransporter pathway. *Proceedings of the National Academy of Sciences* **110**, E4246-E4255 (2013).

Publications

OPEN ACCESS Freely available online

PLOS | PATHOGENS

Lectin-Like Bacteriocins from *Pseudomonas* spp. Utilise D-Rhamnose Containing Lipopolysaccharide as a Cellular Receptor

Laura C. McCaughey[1☯], Rhys Grinter[1☯], Inokentijs Josts[1], Aleksander W. Roszak[2,3], Kai I. Waløen[1,4], Richard J. Cogdell[3], Joel Milner[3], Tom Evans[1], Sharon Kelly[3], Nicholas P. Tucker[5], Olwyn Byron[4], Brian Smith[3], Daniel Walker[1]*

1 Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom, **2** WestCHEM, School of Chemistry, College of Science and Engineering, University of Glasgow, Glasgow, United Kingdom, **3** Institute of Molecular Cell and Systems Biology, College of Medical, Veterinary, and Life Sciences, University of Glasgow, Glasgow, United Kingdom, **4** School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom, **5** Strathclyde Institute for Pharmaceutical and Biomedical Sciences, University of Strathclyde, Glasgow, United Kingdom

Abstract

Lectin-like bacteriocins consist of tandem monocot mannose-binding domains and display a genus-specific killing activity. Here we show that pyocin L1, a novel member of this family from *Pseudomonas aeruginosa*, targets susceptible strains of this species through recognition of the common polysaccharide antigen (CPA) of *P. aeruginosa* lipopolysaccharide that is predominantly a homopolymer of D-rhamnose. Structural and biophysical analyses show that recognition of CPA occurs through the C-terminal carbohydrate-binding domain of pyocin L1 and that this interaction is a prerequisite for bactericidal activity. Further to this, we show that the previously described lectin-like bacteriocin putidacin L1 shows a similar carbohydrate-binding specificity, indicating that oligosaccharides containing D-rhamnose and not D-mannose, as was previously thought, are the physiologically relevant ligands for this group of bacteriocins. The widespread inclusion of D-rhamnose in the lipopolysaccharide of members of the genus *Pseudomonas* explains the unusual genus-specific activity of the lectin-like bacteriocins.

Citation: McCaughey LC, Grinter R, Josts I, Roszak AW, Waløen KI, et al. (2014) Lectin-Like Bacteriocins from *Pseudomonas* spp. Utilise D-Rhamnose Containing Lipopolysaccharide as a Cellular Receptor. PLoS Pathog 10(2): e1003898. doi:10.1371/journal.ppat.1003898

Editor: Guy Tran Van Nhieu, Collège de France, France

Received July 18, 2013; **Accepted** December 10, 2013; **Published** February 6, 2014

Funding: LCM and IJ are supported by 4-year studentships from the Wellcome Trust. RG is supported by a Kelvin Smith Scholarship from the University of Glasgow. We would like to acknowledge the Diamond Light Source for access to I04, I04-1 and I24 (proposal number MX6683). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Daniel.Walker@glasgow.ac.uk

☯ These authors contributed equally to this work.

Introduction

The ability to target a subgroup of pathogenic bacteria in a complex bacterial community has potential applications in medicine and agriculture where the maintenance of a 'normal' microbiome is beneficial. For example, the use of broad spectrum antibiotics to treat bacterial infections is known to cause a range of complications associated with collateral damage to the microbiome, including antibiotic associated diarrhea and *Clostridium difficile* infection [1,2]. In addition, there is growing evidence to suggest that microbial dysbiosis may play a role in a range of chronic diseases such as inflammatory bowel disease, diabetes, obesity and rheumatoid arthritis [3,4,5,6]. Indeed, for Crohn's disease, where the link with dysbiosis is well established, the administration of multiple courses of antibiotics is associated with an increased risk factor for the development of this chronic form of inflammatory bowel disease [7,8,9].

In contrast to the broad spectrum antibiotics that are widely used in medicine and agriculture, protein antibiotics known as bacteriocins often target a specific bacterial species or a group of closely related bacterial species [10,11,12,13]. Well characterised bacteriocins include the S-type pyocins from *P. aeruginosa* and the closely related colicins of *E. coli* [12,13]. The colicin-like bacteriocins form a diverse family of multidomain protein antibiotics which share similar mechanisms of uptake and kill cells through either a pore-forming activity, a specific nuclease activity against DNA, tRNA or rRNA or through inhibition of cell wall synthesis [14,15,16,17]. In the case of S-type pyocins it is thought that their activity is limited to strains of *P. aeruginosa*, whereas colicins show activity against *E. coli* and some strains of closely related bacteria such as *Salmonella* spp. [18]. In the case of colicins and S-type pyocins, killing specificity is primarily determined by the presence of a specific outer membrane receptor on the cell surface. For example, the well characterised E group colicins utilise the TonB-dependent BtuB receptor, which has a normal physiological role in vitamin B_{12} uptake [19]. Colicin-like bacteriocins have also been shown to have a potent antibiofilm activity, indicating their potential as useful therapeutics for the treatment of chronic biofilm mediated infections [20,21]. In the case of the opportunistic human pathogen *P. aeruginosa* there is an urgent requirement for the development of novel therapeutic options since its ability to form drug-resistant biofilms in

Author Summary

Due to rapidly increasing rates of antibiotic resistance observed among Gram-negative pathogens, such as *Pseudomonas aeruginosa*, there is an urgent requirement for novel approaches to the treatment of bacterial infections. Lectin-like bacteriocins are highly potent protein antibiotics that display an unusual ability to kill a select group of bacteria within a specific genus. In this work, we show how the lectin-like protein antibiotic, pyocin L1, can kill *Pseudomonas aeruginosa* with extraordinary potency through specific binding to the common polysaccharide antigen (CPA) of *P. aeruginosa* lipopolysaccharide. The CPA is predominantly a homopolymer of the sugar D-rhamnose that although generally rare in nature is found frequently as a component of the lipopolysaccharide of members of the genus *Pseudomonas*. The targeting of D-rhamnose containing polysaccharides by pyocin L1 and a related lectin-like protein antibiotic, putidacin L1, explains the unusual genus- specific killing activity of the lectin-like bacteriocins. As we learn more about the link between changes to the microbiome and a range of chronic diseases there is a growing realisation that the ability to target specific bacterial pathogens while maintaining the normal gut flora is a desirable property for next generation antibiotics.

combination with the presence of an outer membrane that is highly impermeable to many classes of antibiotics can make this pathogen essentially untreatable in some groups of patients. This is exemplified in cystic fibrosis patients where chronic lung infection with *P. aeruginosa* is the leading cause of mortality [22].

An interesting addition to this group of protein antibiotics is the recently discovered lectin-like bacteriocins that contain two carbohydrate-binding domains of the monocot mannose-binding lectin (MMBL) family [23,24,25,26,27]. Lectin-like bacteriocins from *P. putida* (putidacin L1 or $LlpA_{BW}$) *P. syringae* ($LlpA_{Pss642}$) and *P. fluorescens* ($LlpA1_{Pf\text{-}5}$) have been characterised and have the unprecedented ability to kill strains of a broad range of bacterial species within the genus *Pseudomonas*, but are not active outside this genus [24,26,27]. Similarly the lectin-like bacteriocin $LlpA_{Xcm761}$ from *Xanthomonas citri* pv. *malvacearum* LMG 761 has the ability to kill various species within the genus *Xanthomonas* [24]. The molecular basis of this unusual genus specific activity has not been explained.

Lectins are a structurally and evolutionarily diverse class of proteins produced widely by prokaryotes and eukaryotes and are defined by their ability to recognise and bind carbohydrates. This binding is generally highly specific and mediates a range of diverse functions, including cell-cell interaction, immune recognition and cytotoxicity [28,29] MMBLs represent a structurally conserved lectin subclass, of which the mannose-binding *Galanthus nivalis* agglutinin (GNA) was the first to be characterised [30]. The MMBL-fold consists of a three sided β-prism; each face of which contains a sugar binding motif with the conserved sequence QxDxNxVxY [31]. While originally identified in monocots like *G. nivalis* or *Allium sativum*, it is now recognised that proteins of this class are distributed widely throughout prokaryotes and eukaryotes, where they have evolved to mediate diverse functions [30,32,33,34]. Structural and biochemical analysis of MMBLs has shown that they are generally translated as a single polypeptide chain containing tandem β-prism domains that are then proteolytically processed into monomers. These domains often form homo- or hetero-dimers by strand exchange and π-stacking [35].

The lectin-like bacteriocins are not proteolytically processed and thus consist of a single peptide chain, containing tandem β-prism domains. Sequence alignments of members of this class from *Pseudomonas* spp. show complete conservation of two sugar binding motifs on the C-terminal domain and partial conservation of two sites on the N-terminal domain [23]. Recent work by Ghequire *et al* [23] on the characterisation of putidacin L1 shows these motifs to be important for cytotoxicity. Mutagenesis of the first C-terminal motif has the most dramatic effect on activity, while mutagenesis of the second C-terminal and first N-terminal sugar binding motifs leads to a synergistic reduction in activity. This study also showed low-affinity binding between putidacin L1 and methyl-α-D-mannose or a range of mannose containing oligosaccharides. However, K_ds for these protein-carbohydrate complexes were reported in the range from 46 mM for methyl-α-D-mannoside to 2 mM for a mannose containing pentasaccharide [23]. An extensive search for high affinity carbohydrate binding through the use of glycan arrays failed to detect high affinity carbohydrate binding for this lectin-like bacteriocin [23].

Despite progress in our understanding of the structure and host range of MMBL-like bacteriocins, the mechanism by which these bacteriocins target susceptible strains and exert their antimicrobial effects is unknown. Here we report on the discovery of a novel member of this family, pyocin L1 from *P. aeruginosa*, and show that it utilises lipopolysaccharide (LPS) as a surface receptor, specifically targeting the common polysaccharide antigen (CPA) that is a conserved homopolymer of D-rhamnose. Structural and biophysical analysis shows that the C-terminal carbohydrate binding motifs are responsible for D-rhamnose recognition and that these sites are specific for this sugar over D-mannose. Further to this, we show that the previously described putidacin L1 also selectively binds LPS from susceptible, but not from resistant, *P. syringae* isolates and shows selectivity for D-rhamnose over D-mannose. This work shows that the physiologically relevant ligand for the QxDxNxVxY carbohydrate binding site of the lectin-like bacteriocins is indeed D-rhamnose and not D-mannose as previously thought. As such, the genus-specific activity of lectin-like bacteriocins from *Pseudomonas spp.* can be attributed to the widespread inclusion of the rare D-rhamnose in the LPS of members of the genus *Pseudomonas*.

Results

Identification and characterisation of pyocin L1

As part of a wider project, aimed at identifying bacteriocins that could be used as novel therapeutics in the treatment of *P. aeruginosa* infections, we searched the genomes of 10 recently sequenced clinical and environmental isolates of *P. aeruginosa* for genes with homology to known bacteriocins. One putative bacteriocin gene identified in strain C1433, an isolate from a patient with cystic fibrosis, encodes a protein with 31% identity to the lectin-like bacteriocin $LlpA1_{Pf\text{-}5}$, from *P. fluorescens*. This protein, designated pyocin L1, contained 256-amino acids with a predicted molecular mass of 28413 Da. Alignment of the pyocin L1 protein sequence with other lectin-like bacteriocins, $LlpA1_{Pf\text{-}5}$, $LlpA_{Pss642}$, putidacin L1 ($LlpA_{BW}$), $LlpA_{Au1504}$ from *Burkholderia cenocepacia* and $LlpA_{Xcm761}$ from *Xanthomonas citri* pv. *malvacearum* shows that pyocin L1 contains tandem MMBL domains with three conserved QxDxNxVxY MMBL sugar-binding motifs (Figure S1). Two of these motifs are located in the C-terminal domain of the protein and one in the N-terminal domain. Comparison with the sequences of other lectin-like bacteriocins shows that the C-terminal QxDxNxVxY motifs are highly conserved, with only $LlpA_{Xcm761}$ lacking one C-terminal motif. In contrast the

N-terminal sugar-binding motifs are less well conserved with only $LlpA_{Au1504}$ possessing two fully conserved QxDxNxVxY motifs (Figure S1).

In order to determine the killing spectrum of pyocin L1 we cloned the pyocin L1 open reading frame into the pET21a vector and expressed and purified the protein by nickel affinity, anion exchange and size exclusion chromatography. Purified pyocin L1 was tested for its ability to inhibit the growth of 32 environmental and clinical isolates of *P. aeruginosa* using an overlay spot plate method on LB agar [36]. Under these conditions, pyocin L1 showed killing activity against nine of the *P. aeruginosa* strains tested. Strain E2, an environmental isolate from a tomato plant for which the genome sequence is available, and strain P8, a clinical isolate from a cystic fibrosis patient, showed the greatest sensitivity to pyocin L1 with killing observed down to concentrations of 27 nM and 7 nM, respectively. Pyocin L1 also showed activity against 5 of the 11 *P. syringae* strains tested, although the effect was much weaker, with cell killing observed at high μM concentrations.

Pyocin L1 targets the common polysaccharide antigen (CPA) of *P. aeruginosa* LPS

In order to gain insight into the bacteriocidal activity of pyocin L1, we subjected *P. aeruginosa* E2 to high concentrations of recombinant protein and recovered mutants with greatly increased tolerance to pyocin L1 (Figure 1A). The genomes of two of these mutants were sequenced and comparative analysis with the genome of wild-type E2 revealed a dinuclear deletion, C710 and T711, of the 1146-bp *wbpZ* gene. This deletion was common to both mutants. *wbpZ* encodes a glycosyltransferase of 381 amino acids that plays a key role in lipopolysaccharide synthesis, specifically in the synthesis of the common polysaccharide antigen (CPA) also known as A-band LPS [37], (Figure S2). Most strains of *P. aeruginosa* produce two distinct LPS-types that differ in their O-antigen, but share the same core oligosaccharide. The CPA is predominantly a homopolymer of D-rhamnose and the O-specific antigen contains a heteropolymeric repeating unit that varies widely among strains [38]. Consistent with mutation of *wbpZ*, we found that production of CPA, as determined by immunoblotting with a CPA-specific monoclonal antibody [39], in both M4(E2) and M11(E2) was reduced to undetectable levels (Figure 1B). Visualisation of LPS from these strains was performed via silver staining and comparable quantities of LPS were shown to be present. These observations suggest that pyocin L1 may utilise CPA as a cellular receptor. To test this idea further, we obtained two transposon insertion mutants of *P. aeruginosa* PAO1, which is sensitive to pyocin L1, with insertions in the genes responsible for the transport of CPA to the periplasm [40]. These two genes, *wzt* and *wzm*, encode the ATP-binding component and membrane component of a CPA dedicated ABC transporter [38]. Pyocin L1, which shows good activity against PAO1 showed no activity against strains with insertions in *wzm* and *wzt* (Figure 1C) and immunoblotting with a CPA-specific antibody confirmed the absence of the CPA in these pyocin L1 resistant strains (Figure 1D). Thus, the presence of CPA on the cell surface is required for pyocin L1 killing.

In order to determine if the requirement for CPA is due to a direct interaction with pyocin L1 we purified LPS from wild-type PAO1 and from the pyocin L1 resistant, *wzm* and *wzt* mutants (which produce no CPA but do produce the O-specific antigen) and analysed the pyocin-CPA interaction by isothermal titration calorimetry (ITC). Titration of pyocin L1 into isolated LPS-derived polysaccharides (a mixture of CPA and the O-specific antigen containing polysaccharides) from PAO1 gave rise to strong

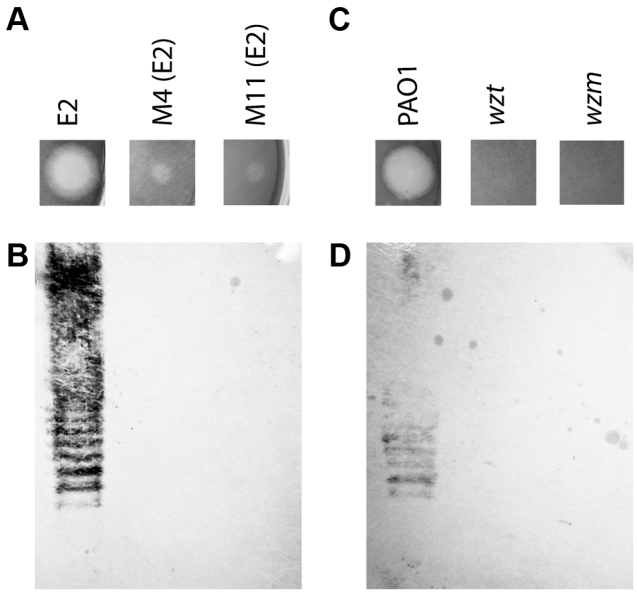


Figure 1. CPA production correlates with pyocin L1 killing. (A) Inhibition of growth of *P. aeruginosa* E2 and tolerant mutants M4 and M11 by pyocin L1, as shown by a soft agar overlay spot-test. 5 μl of purified pyocin L1 (1.5 mg ml^{-1}) was spotted onto a growing lawn of cells. Clear zones indicate cell death. (B) Expression of CPA by *P. aeruginosa* E2 and tolerant mutants, visualised by immunoblotting with the CPA specific antibody N1F10. (C) Inhibition of growth of *P. aeruginosa* PAO1 and PAO1 *wzm* and *wzt* mutants by pyocin L1 (details as for A). (D) Expression of CPA by PAO1 and *wzm* and *wzt* strains (details as for B).
doi:10.1371/journal.ppat.1003898.g001

saturable exothermic heats of binding (Figure 2A), whereas no binding was detected on titration of pyocin L1 into an equivalent concentration of LPS carbohydrates from PAO1 *wzt*, which produces the O-specific antigen but not the CPA (Figure 2B). These data show that pyocin L1 binds directly to the CPA and that this interaction is required for killing. The CPA is therefore likely to be the cellular receptor for pyocin L1.

Pyocin L1 binds the monosaccharide D-rhamnose

The evolutionary relationships between MMBL-like bacteriocins and the originally identified mannose-binding members of this protein family, led to the assumption that carbohydrate binding of polysaccharides by the lectin-like bacteriocins is primarily mediated through binding of D-mannose at one or more of their conserved QxDxNxVxY carbohydrate binding motifs. Indeed, the recent structures [23] of putidacin L1 bound to mannose-containing monosaccharides adds weight to this idea, although measured affinities between polysaccharides and putidacin L1 are weak (mM) and so may not be physiologically relevant. However, the strong interaction between pyocin L1 and CPA, is incompatible with this and suggests that D-rhamnose and not D-mannose is the likely physiological substrate for the QxDxNxVxY carbohydrate binding motifs.

To determine the affinity of pyocin L1 for D-rhamnose and D-mannose, isothermal titration calorimetry (ITC) was performed. Titration of pyocin L1 into D-rhamnose gave rise to weakly saturable heats of binding that are significantly larger than the heats observed on titration of pyocin L1 into an identical concentration of D-mannose (Figure 3). From this experiment an apparent K_d of 5–10 mM was estimated for the interaction of pyocin L1 with D-rhamnose with apparently weaker binding for

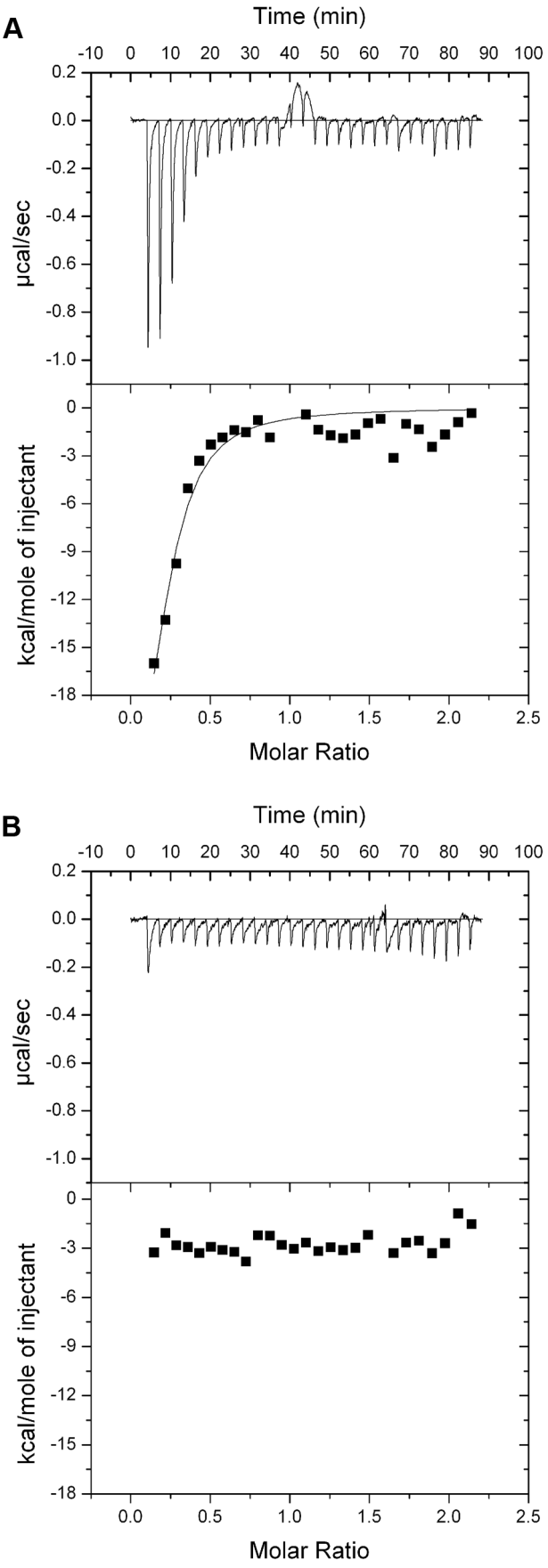


Figure 2. Pyocin L1 binds strongly to CPA from *P. aeruginosa* PAO1. (A) ITC binding isotherm of pyocin L1 (150 µM) titrated into isolated LPS-derived polysaccharide (1 mg ml^{-1}) from wild-type *P. aeruginosa* PAO1. Strong, saturable heats were observed indicative of a strong interaction. Curve fitted with a single binding site model. (B) ITC isotherm of pyocin L1 (150 µM) titrated into isolated LPS-derived polysaccharide (1 mg ml^{-1}) from PAO1 *wzt*. No saturable binding isotherm was observed.
doi:10.1371/journal.ppat.1003898.g002

D-mannose, K_d>50 mM. The interaction between pyocin L1 and these monosaccharides was also probed using NMR with ^{15}N labelled pyocin L1, monitoring changes to its ^{15}N-heteronuclear single quantum correlation (^{15}N-HSQC) spectra on addition of D-rhamnose or D-mannose. In the absence of added monosaccharide ^{15}N-HSQC spectra of pyocin L1, which should contain one crosspeak for each non-proline amide NH as well as peaks for the NH groups in various side chains, were well resolved and dispersed, indicative of a folded protein. Chemical shift perturbation monitored by ^{15}N-HSQC allows the mapping of changes to a protein that occur on ligand binding. Addition of either D-rhamnose or D-mannose up to a concentration of 100 mM did not give rise to large or global changes in chemical shifts (Figure S3). On addition of D-rhamnose significant chemical shift changes were observed for a discrete subset of peaks including some in the amide side chain region of the spectra, while changes of a smaller magnitude were observed on the addition of equal concentrations of D-mannose (Figure S3). Fitting the chemical shift changes that occur on addition of D-rhamnose, for peaks showing strong shifts, to a single site binding model indicates a K_d for the pyocin L1- D-rhamnose complex in the range of 5–20 mM (Figures 3C–F). These data correlated well with the ITC sugar binding data, with low mM binding of pyocin L1 to D-rhamnose and much weaker binding to D-mannose.

D-rhamnose and the CPA bind to the C-terminal QxDxNxVxY motifs of pyocin L1

In an attempt to determine the location of the pyocin L1 D-rhamnose binding site(s) and the structural basis of the D-rhamnose specificity of pyocin L1 we determined the X-ray structures of pyocin L1 with bound D-mannose, D-rhamnose and in the unbound form (Table 1). Pyocin L1, as predicted by sequence homology to MMBL proteins, consists of two tandem β-prism domains characteristic of MMBLs, connected by antiparallel strands propagating from the end of each MMBL domain and lending a strand to the reciprocal β-prism. The strands contain a tryptophan residue which forms π-stacking interactions with two other tryptophans in the β-prism to stabilise the structure (Figure 4A). This interaction is conserved throughout MMBLs, with most members of the class utilising it to form either homo- or hetero-dimers of single MMBL subunits. However, in pyocin L1, as with the recently described structure of putidacin L1, both domains are from a single polypeptide chain [23]. Other structural elements are also common between the two bacteriocins, namely a C-terminal extension of 30 amino acids and a two-turn α-helix insertion into loop 6 of the N-terminal MMBL domain (Figure 4B). The overall root mean square deviation (rmsd) of backbone atoms for pyocin L1 and putidacin L1 is 7.5 Å, which is relatively high due to a difference in the relative orientation of the two MMBL domains. In contrast, the relative orientation of the tandem MMBL domains of pyocin L1 matches those of the dimeric plant lectins very closely, with alignment of pyocin L1 with the snowdrop lectin homodimer (pdb ID: 1MSA) giving an rmsd of 4.81 Å. Comparison of the respective N- and C- terminal domains from pyocin L1 and

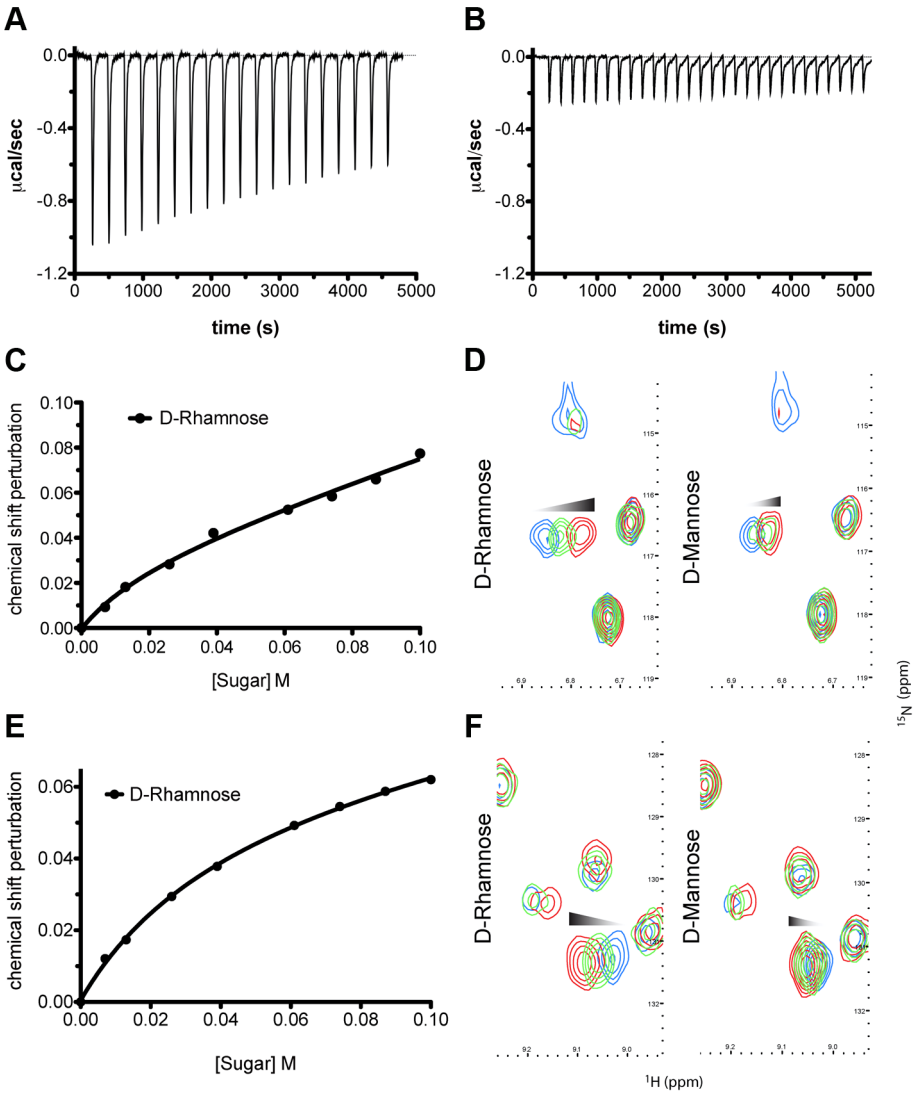


Figure 3. Pyocin L1 shows specificity for D-rhamnose compared with D-mannose. (A) ITC binding isotherm of D-rhamnose (50 mM) titrated into pyocin L1 (100 µM). Weakly saturable heats were observed, indicative of binding with modest affinity (Kd ~5–10 mM). (B) ITC binding isotherm of D-mannose (50 mM) titrated into pyocin L1 (100 µM). Small-weakly saturable heats were observed, indicative of very weak interaction (Kd ~50 mM). Titration of monomeric sugars into ^{15}N-labelled pyocin L1, monitored using ^{1}H-^{15}N HSQC NMR spectroscopy. Shifts within spectra were converted to chemical shift perturbation (CSP) values using equation $\Delta_{ppm} = \sqrt{[\Delta\delta_{HN}+(\Delta\delta_{N}*\alpha_{N})^2]}$. CSP values are plotted against sugar concentration in (C) and (E) and visualised in (D) and (F). Peak positions, which correspond to backbone amide signals, at selected sugar concentrations (blue: no sugar, green: 60 mM, red: 100 mM) are shown. Perturbation of peak position (ppm) is indicative of association between ligand and protein molecules in solution.
doi:10.1371/journal.ppat.1003898.g003

putidacin L1 shows they possess very similar folds with rmsds of 2.77 Å and 2.02 Å, respectively (Figures 4C–D). The higher value for comparison of the N-terminal domains is due to the presence of a 2-strand extension to β-sheet two of the putidacin L1 N-terminal MMBL domain, which is absent from pyocin L1 and other MMBLs. In order to identify protein structures which share a similar fold to pyocin L1 we submitted the structure of the DALI server (http://ekhidna.biocenter.helsinki.fi/dali_server/start). The DALI server searches the protein data bank (PDB) to identify proteins structurally related to the query structure [41]. Significant structural homology was only identified for putidacin L1 and other proteins previously characterised as containing a MMBL fold such as the snowdrop lectin. MMBL dimers of plant origin often form higher order structures, however small angle X-ray scattering of pyocin L1 showed it to be monomeric in solution (Figure S4).

Electron density maps, derived from both D-mannose and D-rhamnose soaked crystals show clear density for sugar moieties in both sites, C1 and C2 (Figure 5). The sugars refined well in these densities at full occupancy, giving B-factors comparable to the surrounding protein side chains. The canonical MMBL hydrogen bonds observed for both D-mannose and D-rhamnose were the same: Gln to O3, Asp to O2, Asn to O2 and Tyr to O4. In addition, O6 of D-mannose forms a hydrogen bond with Tyr169 in C1 and His194 in C2. As D-rhamnose is C6 deoxy D-mannose, it lacks these interactions (Figure 6). The fact that D-mannose forms an additional hydrogen bond is counter-intuitive given that pyocin L1 has a significantly stronger affinity for D-rhamnose, however Val154, Val163 and Ala166 of C1 and Val184 and Ala191 of C2 form a hydrophobic pocket to accommodate the C6-methyl group of D-rhamnose (Figure S5).

Table 1. Crystallographic data collection and refinement statistics.

	Sugar Free Form	D-Rhamnose Soak	D-Mannose Soak
Data collection[a]			
Space group	$C222_1$	$C222_1$	$C222_1$
Cell dimensions, *a*, *b*, *c* (Å)	53.41, 158.40, 147.67	52.99, 160.65, 150.57	53.42, 162.1, 152.5
Resolution (Å)	36.42 - 2.09 (2.14 - 2.09)	54.99 - 2.37 (2.43 - 2.37)	55.53 -2.55 (2.67 - 2.55)
Solvent content (%)	56	55	56
No. of unique observations	37131 (2751)	26242 (1922)	22096 (2901)
Multiplicity	4.8 (4.9)	4.4 (4.5)	5.5 (5.7)
Completeness (%)	99.0 (99.8)	99.1 (99.5)	99.9 (100.0)
R_{merge} (%)	7.2 (59.2)	5.9 (83.0)	7.1 (85.6)
R_{pim} (%)[b]	4.1 (33.0)	3.4 (44.9)	3.3 (39.2)
Mean *I*/sigma (*I*)	14.3 (2.1)	19.0 (2.1)	13.3 (2.3)
Refinement statistics			
R_{work}/R_{free} (%)	17.8/22.2	20.9/25.7	19.4/24.8
No. of non-hydrogen atoms	4505	4178	4138
RMSD of bond lengths (Å)	0.02	0.015	0.013
RMSD of bond angles (°)	1.96	1.63	1.70
No. of waters	344	95	27
Mean/Wilson plot B-value (Å^2)	40.2/33.8	54.2/43.6	65.9/59.1
Ramachandran plot (%)[c]			
Favoured/Allowed/Outliers	97.2/2.2/0.6	97.4/2.2/0.4	96.6/3.0/0.4
PDB identifier	4LE7	4LED	4LEA

[a]Values in parentheses refer to the highest resolution shell.
[b]$R_{pim} = \Sigma_{hkl}[1/(N-1)]^{1/2}\Sigma_i|I_i(hkl) - <I(hkl)>|/\Sigma_{hkl}\Sigma_i I_i(hkl)$.
[c]Percentages of residues in favored/allowed regions calculated by the program RAMPAGE [68].
doi:10.1371/journal.ppat.1003898.t001

Weak density was observed for both sugars at site N1, however given the high concentrations used in the soak and the overall low binding affinity of pyocin L1 for monomeric sugars, it is unlikely that N1 represents a primary binding site for D-rhamnose (Figure S5). The conserved residues in site N2 form interactions with the C-terminal extension of the protein and as such are inaccessible. Weak density was also observed adjacent to the binding site C1 of mol B in both the soaks and in mol A of the D-rhamnose form. This density may correspond to a peripheral binding site utilised in binding to the carbohydrate chain of LPS, as is observed in the structure of putidacin L1 bound to oligosaccharides [23].

To test the idea that the observed binding of D-rhamnose to sites C1 and C2 is reflective of CPA binding and that this binding is critical to pyocin L1 cytotoxicity, we created pyocin L1 variants in which the conserved aspartic acids of the QxDxNxVxY motifs of the C1 and C2 sugar binding sites were mutated to alanine and compared their cytotoxicity and ability to bind the CPA by ITC with the wild-type protein. Titrations with wild-type pyocin L1 and the D150A (C1) and D180A (C2) variants were performed by titrating protein at a concentration of 100 μM into a solution of LPS-derived polysaccharide (1 mg ml^{-1}) from strain PAO1 (Figure 7). Under these conditions we were able to generate binding isotherms that enabled us to accurately determine an apparent K_d of 0.15 (±0.07) μM for the wild-type pyocin L1-CPA complex. For both the D150A (C1) and D180A (C2) variants, affinity for CPA was reduced. For the pyocin L1 D150A-CPA complex a K_d of 1.52 (±0.51) μM was determined, a 10-fold increase in K_d relative to the wild-type pyocin L1-CPA complex. However, CPA binding to the D180A variant was severely weakened and although heats of binding were still observed the K_d for this complex, which could not be accurately determined, is likely >500 μM. We also produced a double mutant in which both D150A and D180A mutations were present. For this double mutant, no binding to CPA was observed by ITC. These data show that both the C1 and C2 sugar binding motifs are required for full CPA binding, but that the C2 binding site is the major CPA binding determinant. The killing activity of these sugar binding motif variants showed a good correlation with their ability to bind the CPA. Both the D150A and D180A variants showed reduced cytotoxicity against PAO1 relative to pyocin L1, with the D150A showing a greater reduction in activity and for the D150A/D180A variant very low levels of cytotoxicity were observed (Figure 7).

Putidacin L1 binds to *P. syringae* LPS and D-rhamnose

Pyocin L1 targets sensitive strains of *P. aeruginosa* through binding to LPS and utilises this as a cell surface receptor. To determine if LPS binding is common to the homologous and previously characterised lectin-like bacteriocin putidacin L1, we purified this protein and determined if the susceptibility of a number of strains of *P. syringae* correlated with the ability of putidacin L1 to bind to LPS-derived carbohydrates from these strains.

From the five strains of *P. syringae* tested, LMG 5456 and LMG 2222 were found to be highly susceptible to putidacin L1 with killing down to concentrations of 0.3 and 7.6 nM respectively. DC3000 and NCPPB 2563 showed complete resistance and LMG 1247 was highly tolerant (killing down to 0.6 μM). Binding of putidacin L1 to the isolated LPS-derived polysaccharides of the

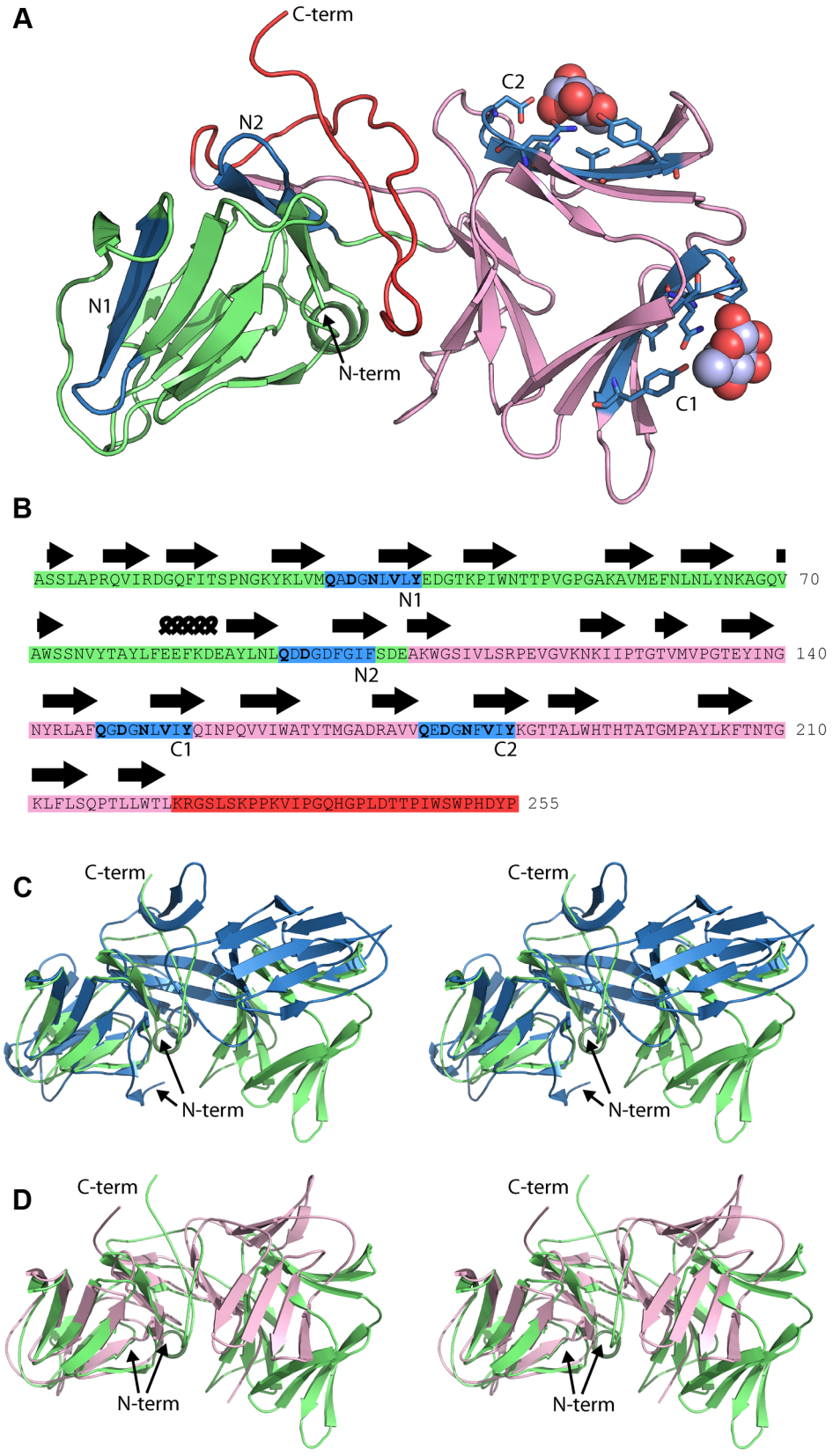


Figure 4. Crystal structure of pyocin L1 reveals tandem MMBL domains and sugar-binding motifs. (A) Ribbon diagram of structure of pyocin L1 in complex with α-D-rhamnose, amino acids 2-256. N-terminal domain (green), C-terminal domain (pink), C-terminal extension (red), α-D-rhamnose (spheres) and sugar binding sites containing the conserved or partially conserved QxDxNxVxY motif are highlighted (blue) and are designated N1, N2 and C1, C2 according to order of appearance in the primary sequence of the N- and C-terminal domains, respectively. Pyocin L1

residues involved in hydrogen bonding with α-D-rhamnose are shown in stick representation. (B) Sequence and secondary structure (β-sheets = arrows, α-helices = coils) of pyocin L1 with colours corresponding to the structure in (A). Residues conserved in sugar binding motifs are shown in bold. (C) Structural alignment of pyocin L1 (green) and putidacin L1 (blue) based on N-terminal MMBL domain in wall-eyed stereo. (D) Structural alignment of pyocin L1 (green) and *Allium sativum* agglutinin (1BWU) (pink) based on N-terminal MMBL domain in wall-eyed stereo.
doi:10.1371/journal.ppat.1003898.g004

above mentioned strains was tested by ITC. Large saturable heats of binding were observed for putidacin L1 and the LPS-derived polysaccharides from LMG 5456 and LMG 2222, while no binding was observed between putidacin L1 and the LPS-derived polysaccharides from LMG 1247, 2563 or DC3000 (Figure 8). Thus, there is excellent correlation between putidacin L1 cell killing and the binding of LPS-derived polysaccharide indicating that like pyocin L1, putidacin L1 utilises LPS as a surface receptor.

Although *P. syringae* O-antigens are diverse relative to CPA, the incorporation of D-rhamnose is widespread and seemingly almost universal in strains of this species [42,43]. Interestingly, in cases where D-rhamnose is not a component of *P. syringae* LPS, L-rhamnose is present [42]. As with pyocin L1 we utilised ITC and NMR to characterise the binding affinity of putidacin L1 for D-rhamnose, in comparison with D-mannose and L-rhamnose. Putidacin L1 exhibited an affinity of 5–10 mM for D-rhamnose, which is comparable to that of pyocin L1, and approximately 10-fold stronger than its affinity for D-mannose (Figure S6). Interestingly, no binding of L-rhamnose to putidacin L1 or pyocin L1 was observed (Figure S7). It is interesting to note that in the strains of *P. syringae* we have tested, the killing spectrum (but not the potency) of pyocin L1 and putidacin L1 is identical. This observation combined with the specificity of putidacin L1 for D-rhamnose, strongly suggests that it also binds to a D-rhamnose containing O-antigen. Indeed branched D-rhamnose O-antigens are common in *P. syringae* [42,43].

Our data for both pyocin L1 and putidacin L1 indicate that D-rhamnose containing O-antigens are utilised as surface receptors for lectin-like bacteriocins from *Pseudomonas* spp. This is an attractive hypothesis since the inclusion of D-rhamnose in the lipopolysaccharides from members of this genus is widespread and

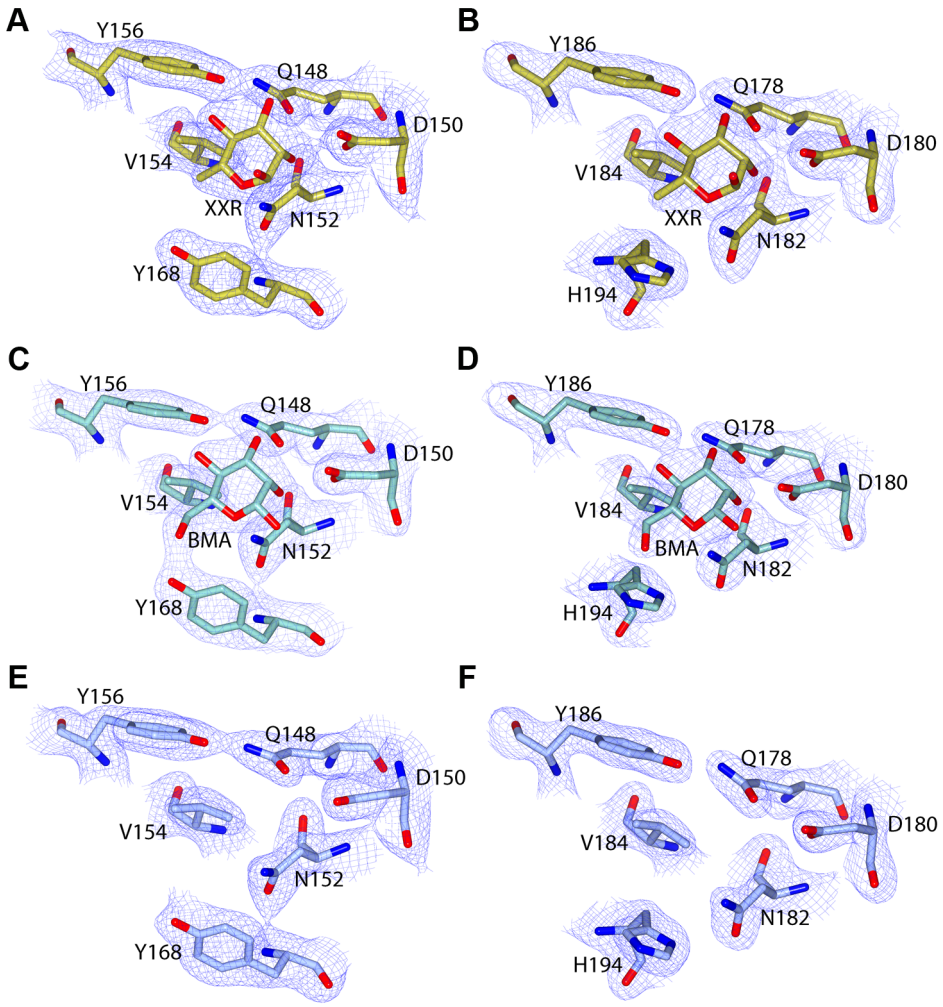


Figure 5. C-terminal MMBL-sugar binding motifs of pyocin L1 bind D-rhamnose and D-mannose. Electron density (at 1.3 σ) with fitted stick model of pyocin L1 MMBL-sugar binding site C1 with: (A) D-rhamnose (XXR), (C) D-mannose (BMA), (E) no bound sugar, and sugar binding site C2 with: (B) D-rhamnose, (D) D-mannose, (F) no bound sugar. For clarity, electron density is clipped to within 1.5 Å of visible atoms.
doi:10.1371/journal.ppat.1003898.g005

could form an important component of the genus specific activity of this group of bacteriocins.

Discussion

In this work we have shown that pyocin L1 targets susceptible cells through binding to the CPA component of LPS and that primary recognition of CPA occurs through binding of D-rhamnose at the conserved QxDxNxVxY sugar binding motifs of the C-terminal lectin domain. The ability of both pyocin L1 and putidacin L1 to recognise D-rhamnose containing carbohydrates is an important component of their ability to target sensitive strains of *Pseudomonas* spp. The use of the O-antigen as a primary receptor differentiates the lectin-like bacteriocins from other multidomain bacteriocins such as colicins and S-type pyocins (colicin-like bacteriocins) which utilise outer membrane proteins as their primary cell surface receptors [44]. The colicin-like bacteriocins also possess a flexible, or natively disordered N-terminal region that is thought to pass through the lumen of a coreceptor and interact with the periplasmic Tol or Ton complexes that mediate translocation of the bacteriocin across the outer membrane [11,44]. The lack of such a flexible N-terminal region in the lectin-like bacteriocins suggests that either they do not need to cross the outer membrane in order to mediate their cytotoxicity or they do so by a mechanism that is fundamentally different to the diverse family of colicin-like bacteriocins. Given the extensive structural homology between the lectin-like bacteriocins and plant lectins it seems likely that these bacteriocins share a common ancestor with plant lectins and from an evolutionary perspective are unrelated to the colicin-like bacteriocins.

In addition to O-antigen recognition, additional factors, as yet to be determined, are clearly also important in strain and species specificity among the lectin-like bacteriocins. Indeed, recent work from Ghequire *et al.* has shown through domain swapping experiments that for putidacin L1 ($LlpA_{BW}$) and the homologous lectin-like bacteriocin $LlpA1_{Pf\text{-}5}$ from *Pseudomonas fluorescens*, species specificity is governed by the identity of the N-terminal lectin domain [23]. Thus, in view of these data and our own data it seems likely that the C-terminal lectin domain of this class of bacteriocins plays a general role in the recognition of D-rhamnose containing O-antigens, with the N-terminal domain interacting with species-specific factors and thus determining the precise species and strain specificity of these bacteriocins. Although there are few clues as to how the lectin-like bacteriocins ultimately kill susceptible cells, we have established a clear role for the C-terminal MMBL domain of these proteins. The roles of the N-terminal MMBL domain and the C-terminal extension remain to be discovered [23]. However, from the previous work of Ghequire *et al*, it is clear that all three of these regions are required for killing of susceptible cells.

Interestingly, although rhamnose is frequently a component of plant and bacterial glycoconjugates, such as the rhamnolipids of *P. aeruginosa* [45] and pectic polysaccharides of plant cell walls [46], it is generally the L-form of this sugar that is found in nature. Although otherwise rare, D-rhamnose is found frequently as a component of the LPS of plant pathogens and plant associated bacteria such as *P. syringae* [42,43], *P. putida* [47], *Xanthomonas campestris* [48] and *Burkholderia* spp. [49], but is a relatively rare component of the O-antigens of animal pathogens such as *E. coli*, *Salmonella* and *Klebsiella*. It is interesting to speculate that since D-rhamnose is a common component of the LPS of bacterial plant pathogens, that some of the many lectins produced by plants may have evolved to target D-rhamnose as part of plant defence to bacterial pathogens.

The specificity of lectin-like bacteriocins suggests that these protein antibiotics may be useful in combating plant pathogenic bacteria, either through the use of bacteriocin expressing biocontrol strains or by the production of transgenic plants engineered to express these proteins. The specific targeting mechanism described here, binding of D-rhamnose containing polymers, indicates that the lectin-like bacteriocins would not interact with either plant or animal cells, since these lack D-rhamnose containing glycoconjugates. In addition, these narrow spectrum antibiotics would leave the majority of the soil microbiome and the gut microbiome of plant-eating animals intact and so would be likely to have minimal environmental impact and minimal impact on animal health. This latter property and the potency of these protein antibiotics could also make the use of lectin-like bacteriocins in the treatment of chronic multidrug-resistant *P. aeruginosa* infections in humans an attractive proposition.

Materials and Methods

Bacterial strains, plasmids and growth conditions

Strains and plasmids utilised in this study are presented in Supplementary Table S1. Strains of *P. aeruginosa* were grown in LB at 37°C, *P. syringae* were grown in King's B Media (KB) (20 g

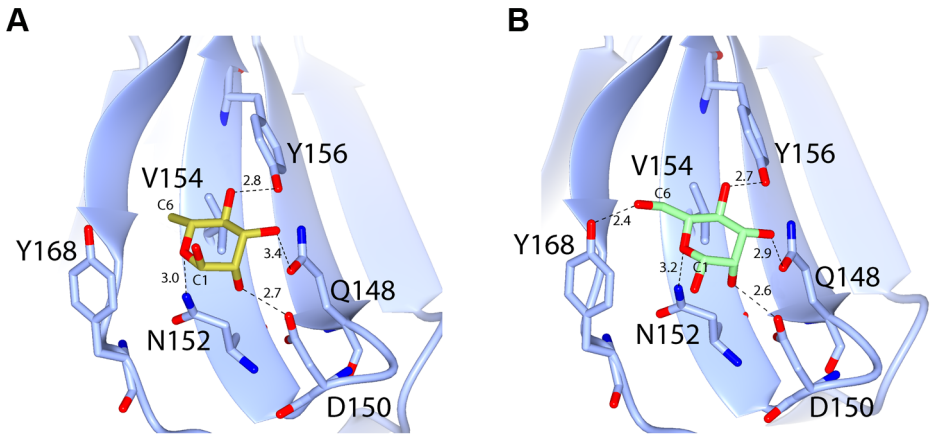


Figure 6. Hydrogen-bonding interactions between pyocin L1 MMBL sugar-binding motif C1 with D-rhamnose and D-mannose. Hydrogen bonds between protein side chains with (A) D-rhamnose and (B) D-mannose are shown; all distances are in Å.
doi:10.1371/journal.ppat.1003898.g006

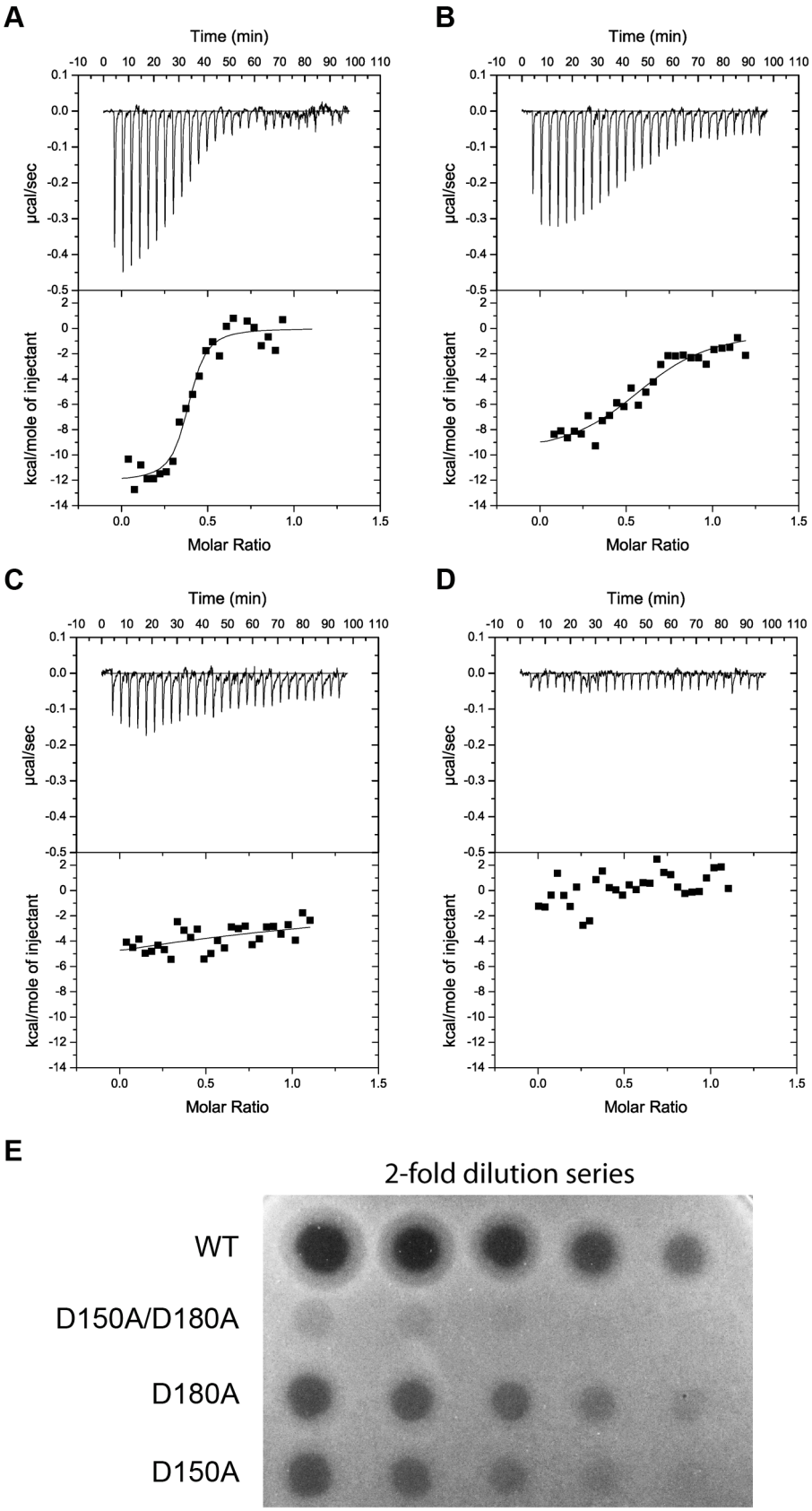


Figure 7. Binding of the CPA at the C-terminal sugar binding motifs, C1 and C2, is critical to pyocin L1 cytotoxicity. ITC binding isotherms of (A) wild-type (B) D180A (C) D150A and (D) D150A/D180A pyocin L1 all at (100 µM) titrated into isolated LPS-derived polysaccharide (1 mg ml^{-1}) from wild-type *P. aeruginosa* PAO1. Fit to a single binding site model is shown. (E) Spot tests to determine cytotoxic activity of wild-type

and pyocin L1 variants against of *P. aeruginosa* PAO1. Purified protein (starting concentration 400 μg ml^{-1} with 2-fold sequential dilutions) was spotted onto a growing lawn of *P. aeruginosa* PAO1. Clear zones indicate pyocin L1 cytotoxicity.
doi:10.1371/journal.ppat.1003898.g007

peptone, 10 g glycerol, 1.5 g $MgSO_4$, 1.5 g K_2HPO_4 per liter adjusted to pH 7.5) at 28°C.

Cloning and purification of lectin-like bacteriocins

Pyocin L1 was amplified from the genomic DNA of the producing strain *P. aeruginosa* C1433 [50] by PCR using primers designed to introduce an NdeI site at the start of the *pyoL1* gene (ACA GAT CAT ATG AAG TCT CCA AAC AAA AGG AGG) and an XhoI site at the end of the gene (ACA GAT CTC GAG GAC CAC GGC GCG CCG TCG TGG ATA GTC GTG GGG CCA A). The PCR product was ligated into the corresponding sites of the *E. coli* expression vector pET21a to give pETPyoL1 which encodes pyocin L1 with a C-terminal His_6 tag separated from the C-terminus of pyocin L1 by a 6 amino acid linker (RRRAVV). Pyocin L1 was overexpressed from *E. coli* BL21(DE3)pLysS carrying the plasmid pETPyoL1. Five litres of LB broth was inoculated (1:100) from an overnight culture and cells were grown at 37°C in a shaking incubator to an $OD_{600} = 0.6$. Protein production was induced by the addition of 0.3 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), the cells were grown at 22°C for a further 20 hand harvested by centrifugation. Cells were resuspended in 20 mM Tris-HCl, 500 mM NaCl, 5 mM imidazole (pH 7.5) and lysed using an MSE Soniprep 150 (Wolf Laboratories) and the cell debris was separated by centrifugation. The cell-free lysate was applied to a 5-ml His Trap HP column (GE Healthcare) equilibrated in 20 mM Tris-HCl, 500 mM NaCl, 5 mM imidazole (pH 7.5) and pyocin L1 was eluted over a 5–500 mM imidazole gradient. Pyocin L1 containing fractions were identified by SDS PAGE, pooled and dialyzed overnight into 50 mM Tris-HCl, 200 mM NaCl, pH 7.5 and remaining contaminants were removed by gel filtration chromatography on a Superdex S75 26/600 column (GE Healthcare) equilibrated in the same buffer. The protein was concentrated using a centrifugal concentrator (Vivaspin 20) with a molecular weight cut off of 5 kDa and stored at −80°C until required. The putidacin L1 open reading frame was synthesised (DNA 2.0) and cloned into pET21a via 5′ NdeI and 3′ XhoI restriction sites. The stop codon was removed in order to utilise the pET21a C-terminal His_6 tag. Purification of putidacin L1 was performed as for pyocin L1. Constructs to express the pyocin L1 mutants D31A, D97A, D150A and D180A were created using the QuikChange Site Directed Mutagenesis Kit (Stratagene) utilising pETPyoL1 as a template. The primers used were CAA ATT GGT CAT GCA AGC GGC TGG CAA CTT GGT CCT TTA CG and CGT AAA GGA CCA AGT TGC CAG CCG CTT GCA TGA CCA ATT TG for D31A, GCG TAC CTG AAT CTT CAA GAT GCT GGG GAC TTC GGT ATA TTT TC and GAA AAT ATA CCG AAG TCC CCA GCA TCT TGA AGA TTC AGG TAC GC for D97A, CGC CTA GCG TTT CAG GGA GCT GGC AAC CTA GTG ATC TAT C and GAT AGA TCA CTA GGT TGC CAG CTC CCT GAA ACG CTA GGC G for D150A and GAT AGA GCA GTA GTG CAA GAG GCT GGA AAT TTT GTT ATC TAC AAA G and CTT TGT AGA TAA CAA AAT TTC CAG CCT CTT GCA CTA CTG CTC

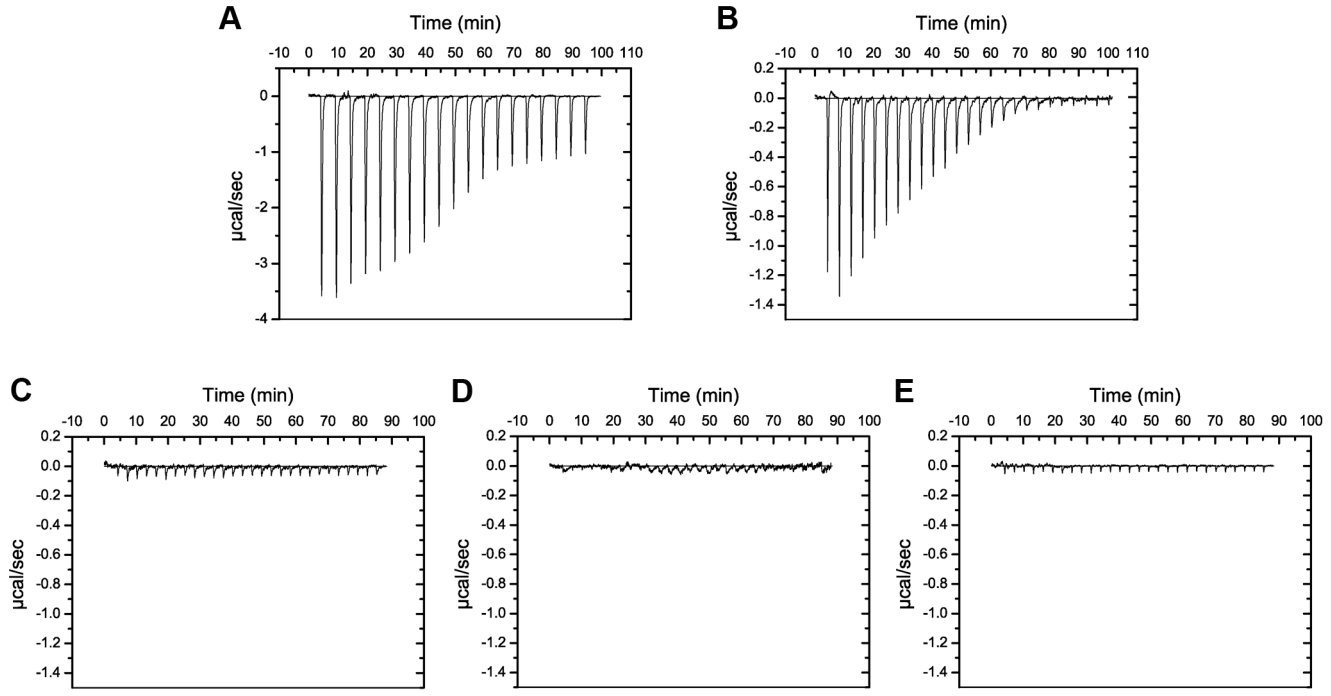


Figure 8. Putidacin L1 binds strongly to LPS-derived polysaccharides from susceptible but not tolerant or resistant *P. syringae* isolates. ITC isotherm of LPS-derived polysaccharides (3 mg ml^{-1}) from strains highly sensitive to putdacin L1: (A) *P. syringae* LMG 2222, (B) *P. syringae* LMG 5456 titrated into putidacin L1 (60 μM). Large, saturable heats are indicative of binding. LPS-derived polysaccharides (3 mg ml^{-1}) from strains non-sensitive to putidacin L1: (C) *P. syringae* NCPPB 2563, (D) *P. syringae* DC3000, or highly tolerant (E) *P. syringae* LMG 1247 to putidacin L1, show no heats of binding when titrated into putidacin L1 (60 μM).
doi:10.1371/journal.ppat.1003898.g008

TAT C for D180A. Mutant proteins were purified as described above for wild-type pyocin L1.

Pyocin sensitivity assays: Overlay spot plate method

Soft agar overlay spot plates were performed using the method of [35]. 150 µl of test strain culture at $OD_{600} = 0.6$ was added to 6 ml of 0.8% soft agar and poured over an LB or KB agar plate. 5 µl of bacteriocin at varying concentrations was spotted onto the plates and incubated for 20 h at 37 or 28°C.

Isolation of pyocin L1 tolerant mutants

1.5 ml of a culture of *P. aeruginosa* E2 ($OD_{600} = 0.6$) was centrifuged and resuspended in 100 µl of LB, to which 100 µl (8 mg ml^{-1}) of purified pyocin L1 was added. The culture was grown for 1 h, plated onto a LB agar plate and incubated for 20 h at 37°C. Isolated colonies were identified as *P. aeruginosa* using 16S PCR as described previously [51].

Whole genome sequencing

The genomes of *P. aeruginosa* E2 and derived pyocin L1 tolerant mutants were sequenced at the Glasgow Polyomics Facility, generating paired-end reads on an Illumina MiSeq Personal Sequencer. Reads were mapped to the previously sequenced parent genomes of *P. aeruginosa* E2 using the CLC genomics workbench, MAUVE and RAST to create an ordered annotated genome. The CLC genomics workbench was used for genome comparisons and the identification of SNPs/INDELs.

LPS purification and isolation of LPS-derived polysaccharide

LPS was purified from 1 litre cultures of *P. aeruginosa* and *P. syringae* strains as described previously, with modifications including the omission of the final trifluoroacetic acid hydrolysis and chromatography steps [52]. Cells were grown for 20 h at 37°C and 28°C for *P. aeruginosa* and *P. syringae* respectively, pelleted by centrifugation at 6000 g for 20 min, and resuspended in 50 mM Tris, pH 7.5 containing lysozyme (2 mg ml^{-1}) and DNase I (0.5 mg ml^{-1}). Cells were lysed by sonication and the cell lysate was incubated at 20°C for 30 min before EDTA was added to a final concentration of 2 mM. An equal volume of aqueous phenol was added and the solution was heated at 70°C for 20 min, with vigorous mixing. The solution was then cooled on ice for 30 min, centrifuged at 7000 g for 20 min and the aqueous phase extracted. Proteinase K was added to a final concentration of 0.05 mg ml^{-1} and dialysed for 12 h against 2×5 L H_2O. LPS was pelleted by ultracentrifugation at 100,000 g for 1 h, resuspended in H_2O and heated to 60°C for 30 min to remove residual proteinase K activity. LPS-derived carbohydrates were isolated by heating LPS in 2% acetic acid for 1.5 h at 96°C. Lipid A was removed by centrifugation at 13,500 g for 3 min followed by extraction with an equal volume of chloroform. The aqueous phase was then lyophilised.

SDS-PAGE, silver staining and immunoblotting

Purified LPS from wild-type and mutant samples were resolved by electrophoresis on 12% SDS-polyacrylamide gels. The LPS banding patterns were visualised by the Invitrogen ultrafast silver staining method. For immunoblotting LPS was transferred onto nitrocellulose membranes and western immunoblotting was performed as previously described using the CPA-specific monoclonal antibody N1F10 and alkaline phosphatase-conjugated goat anti-mouse Fab2 as the secondary antibody [39]. The blots were developed using SIGMAFAS BCIP/NBT tablets.

Isothermal titration calorimetry

ITC experiments were performed on a VP-ITC microcalorimeter (MicroCal LLC). For monosaccharide binding, titrations were carried out at 299 K with regular 15 µl injections of ligands into 60–100 µM pyocin L1 or putidacin L1 at 300 s intervals. 50 mM D-rhamnose, D-mannose or L-rhamnose were used as titrants and reactions were performed in 0.2 M sodium phosphate buffer, pH 7.5. D-rhamnose (>97%) was obtained from Carbosynth Limited (UK) and D-mannose and L-rhamnose (>99%) from Sigma-Aldrich (UK). For O-antigen-pyocin L1 binding reactions, pyocin L1 or pyocin L1 variants were used as titrant at 100 or 150 µM with cleaved O-antigen sugars dissolved at 1 mg ml^{-1} in the chamber. For curve fitting we estimated the molar concentration of LPS-derived CPA containing carbohydrate chains at 20 µM based on an estimated average molecular weight of 10 kDa for CPA containing polysaccharides and estimating the percentage of total LPS represented by CPA containing carbohydrates as 20% of the total by weight [53]. This value may not be accurate and as such the stoichiometry implied by the fit is likely to be unreliable. However, the use of this estimated value has no impact on the reported parameters of ΔH, ΔS and K_d. For O-antigen-putidacin L1 binding reactions, O-antigen was used as the titrant at 3 mg ml^{-1} with 60 µM putidacin L1 in the chamber. Reactions were performed in 20 mM HEPES buffer pH 7.5. All samples were degassed extensively prior to the experiments. Calorimetric data were calculated by integrating the area under each peak and fitted with a single-site binding model with Microcal LLC Origin software. The heats of dilution for each titration were obtained and subtracted from the raw data.

NMR titration experiments

NMR chemical shift perturbation analysis of sugar binding by pyocin L1 and putidacin L1 was carried out at 305 K and 300 K respectively. Fast-HSQC spectra [54] were recorded using ^{15}N labelled proteins (0.1–0.2 mM) and unlabelled ligands, D-rhamnose and D-mannose (100 mM), on a Bruker AVANCE 600 MHz spectrometer. Protein samples were prepared with and without the sugars present and volumes were exchanged at fixed ratios, making sure the protein concentration remained unchanged. The spectra were processed with Topspin and analysed with CCPNmr analysis [55].

Crystallisation and data collection for pyocin L1

Purified pyocin L1 at a concentration of 15 mg ml^{-1} was screened for crystallisation conditions using the Morpheus and PGA crystallisation screens (Molecular Dimensions) [56]. Screens were prepared using a Cartesian Honeybee 8+1 dispensing robot, into 96-well, MRC-format, sitting drop plates (reservoir volume of 80 µl; drop size of 0.5 µl of protein and 0.5 µl of reservoir solution). Clusters of needle shaped crystals grew in a number of conditions in each screen over 3 to 7 days. Two of these conditions, condition 1 (20% v/v ethylene glycol, 10% w/v PEG 8000, 0.03 M $CaCl_2$, 0.03 M $MgCl_2$, 0.1 M Tris/Bicine, pH 8.5) and condition 2 (20% PEG 550 MME, 20% PEG 20 K, 0.03 M $CaCl_2$, 0.03 M $MgCl_2$ 0.1 M MOPS/HEPES, pH 7.5) from the Morpheus screen were selected for optimisation by vapour diffusion in 24 well plates (reservoir volume 500 µl, drop size 1 µl protein and 1 µl reservoir solution). Clusters of needles from these trays grew after 3–7 days and were mechanically separated. The un-soaked crystals were from condition 1, while soaked crystals were from condition 2. Un-soaked crystals were looped and directly cryo-cooled to 110 K in liquid nitrogen; D-mannose and D-rhamnose soaked crystals were soaked for 2–12 min in

artificial mother liquor containing 4 M D-mannose or 2 M D-rhamnose, before cryo-cooling to 110 K. X-ray diffraction data were collected at the Diamond Light Source, Oxfordshire, UK at beam lines I04, I04-1 and I24. Automatic data processing was performed with Xia2 within the EDNA package [57].

Structure solution and refinement for pyocin L1

A dataset from an un-soaked pyocin L1 crystal was submitted to the Balbes pipeline along with the amino acid sequence for pyocin L1 [58]. Balbes produced a partial molecular replacement solution based on the structure of *Galanthus nivalis* agglutinin (PDB ID: 1MSA). Initial phases from Balbes were improved via density modification and an initial model was built using Phase and Build from the Phenix package [59]. The model was then built and refined using REFMAC5 and Coot 0.7 [60,61]. Validation of all models was performed using the Molprobity web server and Procheck from CCP4-I [62,63]. Two structures of sugar soaked pyocin L1 were solved by molecular replacement using Phaser [64], with the sugar-free pyocin L1 as the search model. Additional electron density corresponding to bound sugars, was observed in both $2F_o$-$2F_c$ and F_o-F_c maps [65]. Sugars were fitted and structures refined using Coot 0.7 and REFMAC5. β-D-mannose (PDB ID: BMA) corresponded best to the density of bound D-mannose. The density in the D-rhamnose complex best corresponded to α-D-rhamnose, for which no PDB ligand exists; a model for α-D-rhamnose was prepared by removing the oxygen from carbon 6 of α-D-mannose and submitting these PDB coordinates to the Prodrg server, which generated the model and modeling restraints [65]. The resultant α-D-rhamnose was designated with the PDB ID: XXR.

Small angle X-ray scattering

SAXS was carried out on the X33 beamline at the Deutsches Elektronen Synchrotron (DESY, Hamburg, Germany). Data were collected on samples of Pyocin L1 in the range of 0.5–5 mg ml^{-1}. Buffer was read before and after each sample and an average of the buffer scattering was subtracted from the sample scattering. The data obtained for each sample were analysed using PRIMUS [66], merging scattering data at low angles with high angle data. The distance distribution function, p(r), was obtained by indirect Fourier transform of the scattering intensity using GNOM [67]. A Guinier plot (ln I(s) vs s^2) was used to calculate the molecular weight at I(0) and radius of gyration, R_g, of PyoL1. *Ab initio* models of the protein in solution were built using DAMMIF [68], averaged with DAMAVER [69] and overlaid with the available crystal structure using SUPCOMB [70].

Supporting Information

Figure S1 Sequence alignment of pyocin L1 and previously reported MMBL-like bacteriocins. Dark blue shading designates sequence identity, light blue designates chemically conserved residues. The three conserved MMBL sugar-binding motifs (N1, C1 and C2) and the partially conserved motif (N2) are boxed in red.
(JPG)

Figure S2 Genetics of CPA biosynthesis in *P. aeruginosa*. (A) CPA operon, annotated with location of *P. aeruginosa* E2 tolerant mutant (M4 and M11) deletion and PAO1 transposon insertion mutants. (B) Summary of CPA biosynthetic pathway, showing function performed by genes, shown to induce pyocin L1 tolerance or resistance.
(TIF)

Figure S3 ^{1}H-^{15}N HSQC spectra of ^{15}N-labelled pyocin L1 in presence (red) and absence (black) of 100 mM (A) D-rhamnose and (B) D-mannose, showing distinctive chemical shifts upon addition of associating sugars. Chemical shift changes specific to a small number of cross-peaks illustrates association of the sugars with a small subset of amino acids, which likely correspond to the residues within the binding sites. Analogous changes are observed for D-rhamnose and D-mannose titrations indicative that the same sites are binding both ligands. Greater shift magnitude is observed for D-rhamnose, indicative of a greater affinity towards this monosaccharide. Boxed regions include cross-peaks used for chemical shift perturbation analysis as shown in Figure 3.
(TIF)

Figure S4 Small angle X-ray scattering of pyocin L1. (A) *Ab initio* model of pyocin L1 computed with DAMMIF overlaid with the crystal structure. (B) Guinier plot of scattering data indicates that the protein is monomeric in solution (I(0) gives a molecular mass of 29.53 kDa) by extrapolation of scattering intensity to zero scattering angle. Radius of gyration is 2.72 nm, indicative of a folded, globular monomeric particle in solution.
(TIF)

Figure S5 Coordination of D-rhamnose in C1, C2 and N2 binding sites of pyocin L1. (A) Stereo view of D-rhamnose coordination by binding site C1 (A), C2 (C) and N1 (E), from D-rhamnose soak data. Core binding motif residues (blue) and additional residues contributing to the pocket (white) are shown. Omit map density for D-rhamnose in binding site C1 (B), C2 (D), N1 (F) calculated by refinement of data from D-rhamnose soaked crystal with model built from unsoaked crystal. Density for all sites contoured to 0.15e/$Å^3$.
(TIF)

Figure S6 Putidacin L1 shows specificity for D-rhamnose, compared with D-mannose. (A) ITC isotherm of D-rhamnose (50 mM) titrated into putidacin L1 (0.1 mM). Weakly saturable heats are indicative of binding with modest affinity (Kd ~5–10 mM). (B) ITC isotherm of D-mannose (50 mM) titrated into putidacin L1 (0.1 mM). Binding is undetectable under reaction conditions.
(TIF)

Figure S7 Putidacin L1 and pyocin L1 do not bind L-rhamnose. ITC isotherms of L-rhamnose (50 mM) titrated into putidacin L1 (A) and pyocin L1 (B) both at (0.1 mM). Binding is undetectable under these conditions.
(TIF)

Table S1 Strains and plasmids used in this work.
(PDF)

Text S1 References for supplementary information.
(DOCX)

Acknowledgments

We thank Joseph Lam and Erin Anderson (University of Guelph) for kindly supplying an anti-CPA monoclonal antibody.

Author Contributions

Conceived and designed the experiments: DW LCM RG IJ. Performed the experiments: LCM RG IJ. Analyzed the data: DW LCM RG IJ NPT SK BS KIW AWR OB. Contributed reagents/materials/analysis tools: DW NPT SK BS. Wrote the paper: DW LCM RG IJ AWR KIW RJC JM TE SK NPT OB BS. Provided strains and genome sequences used in the study: NPT.

References

1. Gorkiewicz G (2009) Nosocomial and antibiotic-associated diarrhoea caused by organisms other than Clostridium difficile. International Journal of Antimicrobial Agents 33: S37–S41.
2. Carroll KC, Bartlett JG (2011) Biology of Clostridium difficile: Implications for Epidemiology and Diagnosis. Annual Review of Microbiology 65: 501–521.
3. Manichanh C, Borruel N, Casellas F, Guarner F (2012) The gut microbiota in IBD. Nature Reviews Gastroenterology & Hepatology 9: 599–608.
4. Qin J, Li Y, Cai Z, Li S, Zhu J, et al. (2012) A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 490: 55–60.
5. Henao-Mejia J, Elinav E, Jin C, Hao L, Mehal WZ, et al. (2012) Inflammasome-mediated dysbiosis regulates progression of NAFLD and obesity. Nature 482: 179–U167.
6. Scher JU, Abramson SB (2011) The microbiome and rheumatoid arthritis. Nature Reviews Rheumatology 7: 569–578.
7. Hviid A, Svanstrom H, Frisch M (2011) Antibiotic use and inflammatory bowel diseases in childhood. Gut 60: 49–54.
8. Shaw SY, Blanchard JF, Bernstein CN (2011) Association Between the Use of Antibiotics and New Diagnoses of Crohn's Disease and Ulcerative Colitis. American Journal of Gastroenterology 106: 2133–2142.
9. Spehlmann ME, Begun AZ, Saroglou E, Hinrichs F, Tiemann U, et al. (2012) Risk factors in German twins with inflammatory bowel disease: Results of a questionnaire-based survey. Journal of Crohns & Colitis 6: 29–42.
10. Grinter R, Milner J, Walker D (2012) Ferredoxin containing bacteriocins suggest a novel mechanism of iron uptake in *Pectobacterium* spp. PLoS ONE 7: e33033.
11. Grinter R, Roszak AW, Cogdell RJ, Milner JJ, Walker D (2012) The Crystal Structure of the Lipid II-degrading Bacteriocin Syringacin M Suggests Unexpected Evolutionary Relationships between Colicin M-like Bacteriocins. Journal of Biological Chemistry 287: 38876–38888.
12. Cascales E, Buchanan SK, Duché D, Kleanthous C, Lloubès R, et al. (2007) Colicin biology. Microbiology and Molecular Biology Reviews 71: 158–229.
13. Michel-Briand Y, Baysse C (2002) The pyocins of *Pseudomonas aeruginosa*. Biochimie 84: 499–510.
14. Walker D, Moshbahi K, Vankemmelbeke M, James R, Kleanthous C (2007) The role of electrostatics in colicin nuclease domain translocation into bacterial cells. Journal of Biological Chemistry 282: 31389–31397.
15. Ogawa T, Tomita K, Ueda T, Watanabe K, Uozumi T, et al. (1999) A cytotoxic ribonuclease targeting specific transfer RNA anticodons. Science 283: 2097–2100.
16. Ng CL, Lang K, Meenan NAG, Sharma A, Kelley AC, et al. (2010) Structural basis for 16S ribosomal RNA cleavage by the cytotoxic domain of colicin E3. Nature Structural & Molecular Biology 17: 1241–+.
17. Zeth K, Roemer C, Patzer SI, Braun V (2008) Crystal structure of colicin M, a novel phosphatase specifically imported by Escherichia coli. Journal of Biological Chemistry 283: 25324–25331.
18. Graham AC, Stocker BAD (1977) GENETICS OF SENSITIVITY OF SALMONELLA SPECIES TO COLICIN-M AND BACTERIOPHAGES T5 T1, AND ES18. Journal of Bacteriology 130: 1214–1223.
19. Kurisu G, Zakharov SD, Zhalnina MV, Bano S, Eroukova VY, et al. (2003) The structure of BtuB with bound colicin E3 R-domain implies a translocon. Nature Structural Biology 10: 948–954.
20. Smith K, Martin L, Rinaldi A, Rajendran R, Ramage G, et al. (2012) Activity of Pyocin S2 against Pseudomonas aeruginosa Biofilms. Antimicrobial Agents and Chemotherapy 56: 1599–1601.
21. Brown CL, Smith K, McCaughey L, Walker D (2012) Colicin-like bacteriocins as novel therapeutic agents for the treatment of chronic biofilm-mediated infection. Biochemical Society Transactions 40: 1549–1552.
22. Lyczak JB, Cannon CL, Pier GB (2002) Lung Infections Associated with Cystic Fibrosis. Clinical Microbiology Reviews 15: 194–222.
23. Ghequire MGK, Garcia-Pino A, Lebbe EKM, Spaepen S, Loris R, et al. (2013) Structural Determinants for Activity and Specificity of the Bacterial Toxin LlpA. PLoS pathogens 9: e1003199–e1003199.
24. Ghequire MGK, Li W, Proost P, Loris R, De Mot R (2012) Plant lectin-like antibacterial proteins from phytopathogens Pseudomonas syringae and Xanthomonas citri. Environmental Microbiology Reports 4: 373–380.
25. Ghequire MGK, Loris R, De Mot R (2012) MMBL proteins: from lectin to bacteriocin. Biochemical Society Transactions 40: 1553–U1433.
26. Parret AHA, Schoofs G, Proost P, De Mot R (2003) Plant lectin-like bacteriocin from a rhizosphere-colonizing Pseudomonas isolate. Journal of Bacteriology 185: 897–908.
27. Parret AHA, Temmerman K, De Mot R (2005) Novel lectin-like bacteriocins of biocontrol strain Pseudomonas fluorescens Pf-5. Applied and Environmental Microbiology 71: 5197–5207.
28. Sharon N (2001) Lectins. eLS: John Wiley & Sons, Ltd.
29. Sharon N, Lis H (2004) History of lectins: from hemagglutinins to biological recognition molecules. Glycobiology 14: 53R–62R.
30. Van Damme EJM, Nakamura-Tsuruta S, Smith DF, Ongenaert M, Winter HC, et al. (2007) Phylogenetic and specificity studies of two-domain GNA-related lectins: generation of multispecificity through domain duplication and divergent evolution. Biochemical Journal 404: 51–61.
31. Chandra NR, Ramachandraiah G, Bachhawat K, Dam TK, Surolia A, et al. (1999) Crystal structure of a dimeric mannose-specific agglutinin from garlic: Quaternary association and carbohydrate specificity. Journal of Molecular Biology 285: 1157–1168.
32. Vasta GR, Nita-Lazar M, Giomarelli B, Ahmed H, Du S, et al. (2011) Structural and functional diversity of the lectin repertoire in teleost fish: Relevance to innate and adaptive immunity. Developmental and Comparative Immunology 35: 1388–1399.
33. Kurimoto E, Suzuki M, Amemiya E, Yamaguchi Y, Nirasawa S, et al. (2007) Curculin Exhibits Sweet-tasting and Taste-modifying Activities through Its Distinct Molecular Surfaces. Journal of Biological Chemistry 282: 33252–33256.
34. Shimokawa M, Fukudome A, Yamashita R, Minami Y, Yagi F, et al. (2012) Characterization and cloning of GNA-like lectin from the mushroom Marasmius oreades. Glycoconjugate Journal 29: 457–465.
35. Hester G, Wright CS (1996) The Mannose-specific bulb lectin from Galanthus nivalis (snowdrop) binds mono- and dimannosides at distinct sites. Structure analysis of refined complexes at 2.3 angstrom and 3.0 angstrom resolution. Journal of Molecular Biology 262: 516–531.
36. Fyfe JAM, Harris G, Govan JRW (1984) Revised Pyocin Typing Method For Pseudomonas-Aeruginosa. Journal of Clinical Microbiology 20: 47–50.
37. Rocchetta HL, Burrows LL, Pacan JC, Lam JS (1998) Three rhamnosyltransferases responsible for assembly of the A-band D-rhamnan polysaccharide in Pseudomonas aeruginosa: a fourth transferase, WbpL, is required for the initiation of both A-band and B-band lipopolysaccharide synthesis. Molecular Microbiology 30: 1131–1131.
38. Lam JS, Taylor VL, Islam ST, Hao Y, Kocincova D (2011) Genetic and Functional Diversity of Pseudomonas aeruginosa Lipopolysaccharide. Frontiers in microbiology 2: 118–118.
39. Hao Y, King JD, Huszczynski S, Kocincova D, Lam JS (2013) Five New Genes Are Important for Common Polysaccharide Antigen Biosynthesis in Pseudomonas aeruginosa. Mbio 4.
40. Jacobs MA, Alwood A, Thaipisuttikul I, Spencer D, Haugen E, et al. (2003) Comprehensive transposon mutant library of Pseudomonas aeruginosa. Proceedings of the National Academy of Sciences of the United States of America 100: 14339–14344.
41. Holm L, Rosenström P (2010) Dali server: conservation mapping in 3D. Nucleic Acids Research 38: W545–W549.
42. Ovod V, Rudolph K, Knirel Y, Krohn K (1996) Immunochemical characterization of O polysaccharides composing the alpha-D-rhamnose backbone of lipopolysaccharide of Pseudomonas syringae and classification of bacteria into serogroups O1 and O2 with monoclonal antibodies. Journal of Bacteriology 178: 6459–6465.
43. Ovod VV, Knirel YA, Samson R, Krohn KJ (1999) Immunochemical characterization and taxonomic evaluation of the O polysaccharides of the lipopolysaccharides of Pseudomonas syringae serogroup O1 strains. Journal of Bacteriology 181: 6937–6947.
44. Kleanthous C (2010) Swimming against the tide: progress and challenges in our understanding of colicin translocation. Nature Reviews Microbiology 8: 843–848.
45. Abdel-Mawgoud AM, Lepine F, Deziel E (2010) Rhamnolipids: diversity of structures, microbial origins and roles. Applied Microbiology and Biotechnology 86: 1323–1336.
46. Caffall KH, Mohnen D (2009) The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. Carbohydrate Research 344: 1879–1900.
47. Knirel YA, Shashkov AS, Senchenkova S, Ajiki Y, Fukuoka S (2002) Structure of the O-polysaccharide of Pseudomonas putida FERM p-18867. Carbohydrate Research 337: 1589–1591.
48. Molinaro A, Silipo A, Lanzetta R, Newman MA, Dow JM, et al. (2003) Structural elucidation of the O-chain of the lipopolysaccharide from Xanthomonas campestris strain 8004. Carbohydrate Research 338: 277–281.
49. Vinion-Dubiel AD, Goldberg JB (2003) Lipopolysaccharide of Burkholderia cepacia complex. Journal of Endotoxin Research 9: 201–213.
50. Stewart L, Ford A, Sangal V, Jeukens J, Boyle B, et al. (2013) Draft genomes of twelve host adapted and environmental isolates of Pseudomonas aeruginosa and their position in the core genome phylogeny. Pathogens and Disease [epub ahead of print]
51. Claesson MJ, Wang Q, O'Sullivan O, Greene-Diniz R, Cole JR, et al. (2010) Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. Nucleic Acids Research 38.
52. Ramm M, Lobe M, Hamburger M (2003) A simple method for preparation of D-rhamnose. Carbohydrate Research 338: 109–112.
53. Rivera M, Bryan LE, Hancock REW, McGroarty EJ (1988) Heterogeneity Of Lipopolysaccharides From Pseudomonas-Aeruginosa - Analysis Of Lipopolysaccharide Chain-Length. Journal of Bacteriology 170: 512–521.
54. Mori S, Abeygunawardana C, Johnson MO, vanZijl P (1996) Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation (vol 108, pg 94, 1995). Journal of Magnetic Resonance Series B 110: 321–321.
55. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, et al. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. Proteins-Structure Function and Bioinformatics 59: 687–696.

56. Gorrec F (2009) The MORPHEUS protein crystallization screen. Journal of Applied Crystallography 42: 1035–1042.
57. Incardona M-F, Bourenkov GP, Levik K, Pieritz RA, Popov AN, et al. (2009) EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. Journal of Synchrotron Radiation 16: 872–879.
58. Long F, Vagin AA, Young P, Murshudov GN (2008) BALBES: a molecular-replacement pipeline. Acta Crystallographica Section D-Biological Crystallography 64: 125–132.
59. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D-Biological Crystallography 66: 213–221.
60. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Crystallographica Section D-Biological Crystallography 66: 486–501.
61. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, et al. (2011) REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallographica Section D-Biological Crystallography 67: 355–367.
62. Chen VB, Arendall WB, III, Headd JJ, Keedy DA, Immormino RM, et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallographica Section D-Biological Crystallography 66: 12–21.
63. Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) PROCHECK - A program to check the stereochemical quality of protein structures. Journal of Applied Crystallography 26: 283–291.
64. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) Phaser crystallographic software. Journal of Applied Crystallography 40: 658–674.
65. Schuttelkopf AW, van Aalten DMF (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallographica Section D-Biological Crystallography 60: 1355–1363.
66. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. Journal of Applied Crystallography 36: 1277–1282.
67. Svergun DI (1992) Determination Of The Regularization Parameter In Indirect-Transform Methods Using Perceptual Criteria. Journal of Applied Crystallography 25: 495–503.
68. Franke D, Svergun DI (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. Journal of Applied Crystallography 42: 342–346.
69. Volkov VV, Svergun DI (2003) Uniqueness of ab initio shape determination in small-angle scattering. Journal of Applied Crystallography 36: 860–864.
70. Kozin MB, Svergun DI (2001) Automated matching of high- and low-resolution structural models. Journal of Applied Crystallography 34: 33–41.

Molecular Microbiology (2014) 93(2), 234–246 ■
doi:10.1111/mmi.12655
First published online 18 June 2014

Structure of the atypical bacteriocin pectocin M2 implies a novel mechanism of protein uptake

Rhys Grinter,[1†] Inokentijs Josts,[1†] Kornelius Zeth,[2,3] Aleksander W. Roszak,[4] Laura C. McCaughey,[1] Richard J. Cogdell,[5] Joel J. Milner,[6] Sharon M. Kelly,[5] Olwyn Byron[6] and Daniel Walker[1*]

[1]*Institute of Infection, Immunity and Inflammation*, [5]*Institute of Molecular Cell and Systems Biology and* [6]*School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK.*

[4]*WestCHEM, School of Chemistry, College of Science and Engineering, University of Glasgow, Glasgow G12 8QQ, UK.*

[2]*Unidad de Biofisica (CSIC-UPV/EHU), Barrio Sarriena s/n, 48940, Leioa, Vizcaya, Spain.*

[3]*IKERBASQUE, Basque Foundation for Science, Bilbao, Spain.*

Summary

The colicin-like bacteriocins are potent protein antibiotics that have evolved to efficiently cross the outer membrane of Gram-negative bacteria by parasitizing nutrient uptake systems. We have structurally characterized the colicin M-like bacteriocin, pectocin M2, which is active against strains of *Pectobacterium* spp. This unusual bacteriocin lacks the intrinsically unstructured translocation domain that usually mediates translocation of these bacteriocins across the outer membrane, containing only a single globular ferredoxin domain connected to its cytotoxic domain by a flexible α-helix, which allows it to adopt two distinct conformations in solution. The ferredoxin domain of pectocin M2 is homologous to plant ferredoxins and allows pectocin M2 to parasitize a system utilized by *Pectobacterium* to obtain iron during infection of plants. Furthermore, we identify a novel ferredoxin-containing bacteriocin pectocin P, which possesses a cytotoxic domain homologous to lysozyme, illustrating that the ferredoxin domain acts as a generic delivery module for cytotoxic domains in *Pectobacterium*.

Accepted 23 May, 2014. *For correspondence. E-mail Daniel.Walker@glasgow.ac.uk; Tel. (+44) 141 3305082; Fax (+44) 141 3304297. †These authors contributed equally.

Introduction

It is a dogma of colicin biology that after binding tightly to their cognate outer membrane (OM) receptor, colicins utilize an intrinsically unstructured translocation domain (IUTD) to recruit the inner membrane-bound Tol or Ton complex (Kleanthous, 2010; Housden *et al.*, 2013). These complexes, which are responsive to the proton motive force (pmf), mediate translocation of the bacteriocin across the OM (Cascales *et al.*, 2007; Housden *et al.*, 2010). The formation of a colicin translocon has recently been visualized directly for the DNase-type colicin E9 through the isolation and imaging of the colicin in complex with its primary receptor BtuB, the trimeric porin OmpF, which allows passage of the IUTD across the OM and the periplasmic protein TolB, which is a component of the cell envelope-spanning TolABQR-Pal complex (Housden *et al.*, 2013). Similarly, the TonB-dependent pore-forming colicin, colicin IA, uses one copy of the TonB-dependent receptor Cir as its primary receptor and a second copy as a translocation pathway for its IUTD to cross the OM to deliver a TonB-binding epitope to the periplasm (Jakes and Finkelstein, 2010). In addition to the colicins, which show a potent narrow spectrum of killing activity against strains of *E. coli* and other closely related bacteria, other colicin-like bacteriocins have also been characterized. These include the S-type pyocins from *Pseudomonas aeruginos*a, klebicins from *Klebsiella pneumonia* and syringacin M from *P. syringae* (Riley *et al.*, 2001; Michel-Briand and Baysse, 2002; Barreteau *et al.*, 2009). The recently determined structures of the M-class bacteriocins pyocin M and syringacin M showed that like colicin M, these bacteriocins possess a 30- to 40-amino-acid IUTD, which is essential for translocation, indicating that translocation across the OM likely occurs through the same mechanism as the colicins (Zeth *et al.*, 2008; Barreteau *et al.*, 2012a,b; Grinter *et al.*, 2012b).

We recently described the novel M-class bacteriocins pectocin M1 and M2, which are produced by and active against strains of the soft-rot phytopathogens *Pectobacte-*

rium atrosepticum (Pba) and *Pectobacterium carotovorum (Pbc)* (Grinter *et al*., 2012a; 2013). The domain structure of these proteins suggested that they challenge the dogma that an IUTD is the universal mechanism by which colicin-like bacteriocins achieve translocation. Pectocin M1 and M2 consist of an M-class cytotoxic domain with lipid II degrading activity, fused to a plant-like ferredoxin domain (Grinter *et al*., 2012a; 2013). This ferredoxin domain, which contains an intact [2Fe-2S] iron–sulphur cluster, substitutes for the helical receptor binding domain and IUTD of the M-class bacteriocins discussed above, that are required to deliver the cytotoxic domain to the periplasm. Further to this, we observed that the addition of plant ferredoxin to strains of *Pba* and *Pbc* exposed to the pectocins inhibited bacteriocin-induced killing (Grinter *et al*., 2012a). These observations show first, that *Pba* and *Pbc* possess an OM receptor able to bind ferredoxin and second, that pectocins M1 and M2 parasitize this receptor to target and ultimately gain entry to susceptible cells. The role of ferredoxin binding for these plant pathogens is apparent under iron-limiting conditions where, in the presence of plant ferredoxin, some strains of *Pectobacterium* spp. show strongly enhanced growth (Grinter *et al*., 2012a). This effect is not observed on addition of the mammalian ferredoxin homologue, adrenodoxin, which also contains a [2Fe-2S] iron–sulphur cluster (even at greatly increased concentrations), indicating a high level of specificity for plant ferredoxin. Similarly, adrenodoxin is not able to rescue cells from pectocin M-induced killing (Grinter *et al*., 2012a). Thus, like other colicin-like bacteriocins, pectocins M1 and M2 parasitize an existing nutrient uptake system to gain entry into target cells. However, for these bacteriocins the mechanism is overt and unprecedented, with the direct utilization of ferredoxin, a protein from which *Pectobacterium* spp. is able to directly acquire iron, as the targeting region of the bacteriocin (Grinter *et al*., 2013).

In order to gain further insight into the mechanism through which pectocins M1 and M2 gain entry into target cells, we have used X-ray crystallography and small angle X-ray scattering along with *in silico* modelling approaches to characterize the structural and dynamic properties of pectocin M2. Our data show that there is a high degree of conformational flexibility between the ferredoxin and colicin M-like cytotoxic domain through movement of a linking helix and definitively show that the protein lacks the flexible IUTD that is characteristic of all other characterized colicin-like bacteriocins. The lack of an IUTD indicates that the ferredoxin-containing pectocins utilize an existing ferredoxin uptake mechanism to cross the OM, without direct interaction with the Tol or Ton complexes in the periplasm. Additionally, we have determined the existence of an additional ferredoxin-containing bacteriocin, pectocin P, which possesses a cytotoxic domain that is a structural homologue of lysozyme, illustrating that ferredoxin can act as a generic module for the delivery of structurally diverse cytotoxic proteins to the periplasm.

Results

The crystal structure of pectocin M2

In initial crystallization trials for pectocin M2, characteristic red-brown crystals of this ferredoxin-containing bacteriocin formed with PEG 3350 and ammonium sulphate as the precipitants. Data from these crystals were collected to 2.3 Å in the space group $P2_1$ and phased using anomalous scattering data from the metal centres of the [2Fe-2S] iron–sulphur cluster. The structure of pectocin M2 revealed an N-terminal domain with the predicted ferredoxin-fold (residues 2–94, in red), separated from the colicin M-like cytotoxic domain (residues 116–271, in blue) by a linker region (residues 95–115, in green) that forms an α-helix (Fig. 1A and B). There is a significant difference in the orientation of the cytotoxic and ferredoxin domains of the two pectocin M2 molecules in the asymmetric unit (ASU) with a root mean square deviation (r.m.s.d.) of 3.4 Å, between main-chain atoms (Fig. S1). The fold of the pectocin M2 ferredoxin domain is identical (r.m.s.d. 0.60 Å) to that of spinach ferredoxin (PDB ID = 1A70) and the C-terminal cytotoxic domain is highly similar to the lipid II-cleaving catalytic domains of colicin M (PDB ID = 2XMX, r.m.s.d. 1.7 Å) (Fig. 1C and D) (Zeth *et al*., 2008). The crystal structure of pectocin M2 adds to a growing body of structural and biochemical data on colicin M-like cytotoxic domains (Zeth *et al*., 2008; Barreteau *et al*., 2010; 2012b; Helbig and Braun, 2011; Grinter *et al*., 2012b). We confirmed the enzymatic activity of pectocin M1 and M2 by a lipid II hydrolysis assay (Fig. S2). In the recently solved structures of pyocin M (PaeM) and syringacin M a divalent metal ion (Ca^{2+} or Mg^{2+}) is co-ordinated by a key catalytic aspartic acid side-chain in conjunction with two backbone carbonyls. Mg^{2+}, Ca^{2+} or Mn^{2+} ions are required for catalytic activity of M-class bacteriocins, and analysis of these proteins has shown co-ordination at this position to be essential for activity (Grinter *et al*., 2012b). In the pectocin M2 structure this key aspartate (D226) adopts an analogous conformation. However, no density for a metal ion is observed in this position, which is occupied by a water molecule (Fig. 2A). The absence of a metal ion is unsurprising given the lack of divalent ions and the high ionic strength of the crystallization conditions.

Comparative analysis of the catalytic domains of colicin M homologues reveals significant variation between the structures. In the structures of pyocin M and syringacin M a key conserved arginine is located distant from other conserved residues creating an open active-site cleft. In contrast, in the pectocin M2 structure this residue (R236) is orientated towards the other key catalytic residues

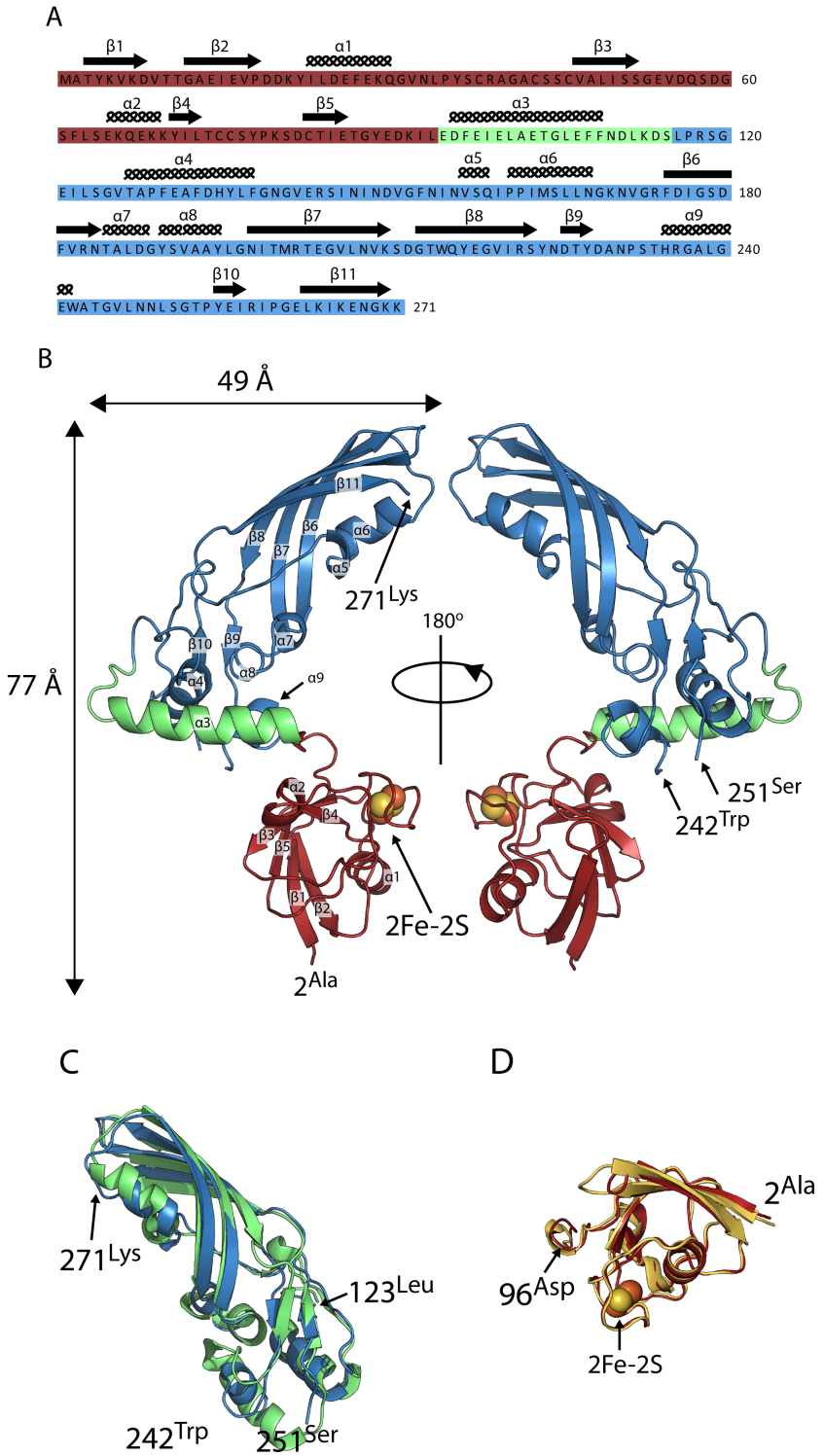


Fig. 1. The crystal structure of pectocin M1 reveals a ferredoxin-containing colicin-like bacteriocin that lacks an IUTD.
A. Amino acid sequence of pectocin M2, showing structural domains (ferredoxin domain = red, linker helix = green, catalytic domain = blue) and annotated with secondary structure.
B. Schematic of the crystal structure of pectocin M2 observed in the $P2_1$ crystal form, with the cytotoxic domain in blue, plant-like ferredoxin domain in red and linker helix in green. The [2Fe-2S] iron–sulphur cluster is represented by spheres.
C. Schematic of cytotoxic domain of pectocin M2 aligned with that of colicin M (PDB ID = 2XMX) (backbone r.m.s.d. = 1.65 Å, pectocin M2 residues = 123–271, colicin M residues = 123–271).
D. Schematic of the ferredoxin domain of pectocin M2 aligned with that of spinach ferredoxin (PDB ID = 1A70) (backbone r.m.s.d. = 0.6 Å, pectocin M2 residues = 2–96, spinach ferredoxin residues 2–96).

(Fig. 2B). The electron density for R236 permitted modelling of two conformations, one within hydrogen bonding distance of the aspartic acid co-ordinated water and the other forming a hydrogen bond with N184. In this conformation, R236 creates a defined active site tunnel which would enable co-ordination of the lipid II and positioning of the pyrophosphate group in close proximity to all key catalytic residues (Fig. 2A and B).

In contrast to the compact structures of the homologous bacteriocins, colicin M, pyocin M and syringacin M, where

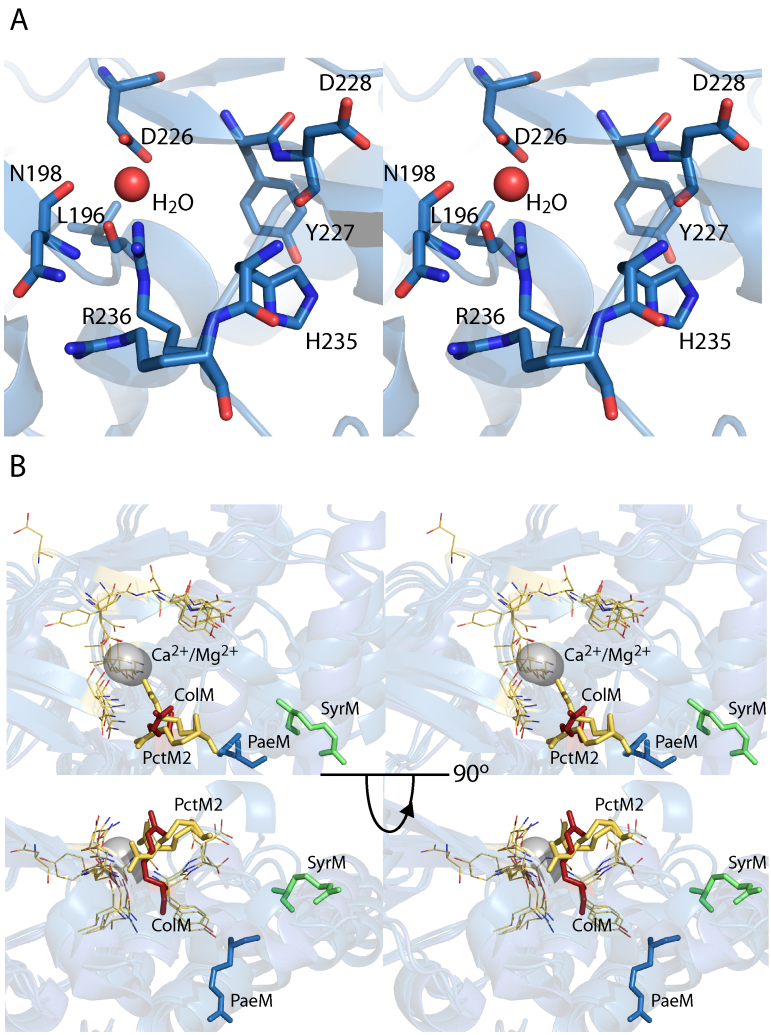


Fig. 2. Colicin M-class bacteriocins possess a highly flexible active site.
A. A stereo view of a stick model of the key active site residues of pectocin M2, showing a water molecule occupying the key metal binding site of the enzyme.
B. A stereo view of the overlay of the catalytic site from all structurally characterized colicin-M class bacteriocins, showing conformational variability of the key catalytic arginine. Key arginine shown as sticks and colour coded according to structure; green = syringacin M (PDB ID = 4FZL), blue = pyocin M (PDB ID = 4G75), red = colicin M (PDB ID = 2XMX) and yellow = pectocin M2 (PDB ID = 4N58), All other catalytically important residues shown as lines in yellow.

the receptor binding and catalytic domains are not separated by linker regions and do not form obviously structurally distinct elements (Zeth *et al.*, 2008; Barreteau *et al.*, 2012b; Grinter *et al.*, 2012b), the catalytic and receptor binding domains of pectocin M2 do not form extensive interactions. The relative orientation of the ferredoxin domain, linker region and cytotoxic domain gives rise to a non-linear dog-leg structure. Interestingly, and again in contrast to colicin M, pyocin M and syringacin M, the N-terminal region of pectocin M2 lacks a disordered or flexible IUTD that is otherwise characteristic of the colicin-like bacteriocins, with the entire N-terminus being integral to the globular ferredoxin domain. These data suggest a mechanism of uptake distinct from closely related colicin-like bacteriocins.

Pectocin M2 is flexible

Given that pectocin M2 lacks an IUTD required to contact the Tol or Ton complexes in the periplasm and mediate translocation of this protein across the outer membrane, alternative mechanisms of uptake must be considered. One possibility is that the entire bacteriocin passes through the lumen of its OM receptor. Since proteins involved in iron uptake are invariably TonB-dependent receptors that possess large 22-stranded β-barrels this may be plausible. However, such a mechanism would only be feasible if pectocin M2 were flexible and significant rearrangement of the dog-leg configuration observed in the crystal structure could be achieved. The observation that there is a relatively large difference in orientation between the cytotoxic and ferredoxin domains in the monomers of the ASU (Fig. S1) is suggestive of such flexibility and indicates that the crystal structure may not be wholly representative of pectocin M2 in solution.

To assess the conformational flexibility of pectocin M2 we performed small angle X-ray scattering (SAXS). SAXS data were obtained for a range of pectocin M2 concentrations. Comparison of these data with a theoretical scattering curve generated, using CRYSOL (Svergun *et al.*,

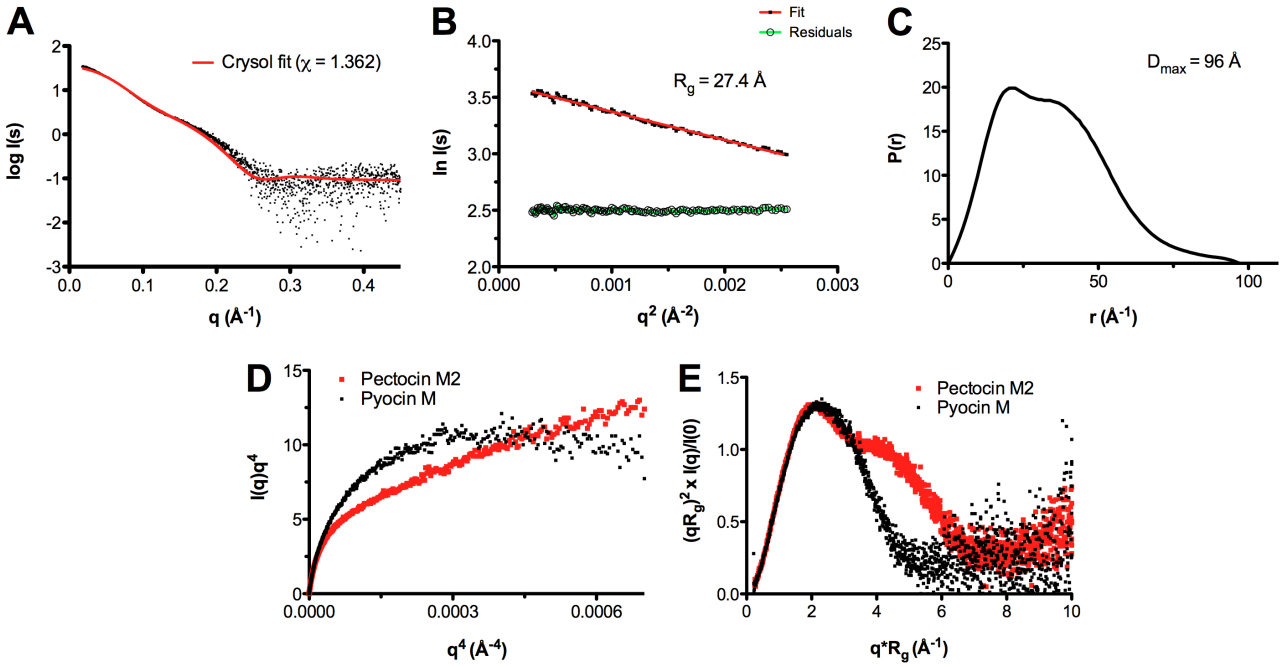


Fig. 3. SAXS shows pectocin M2 is flexible.
A. Overlay of the experimentally determined pectocin M2 SAXS curve (black points) with the scattering curve computed with CRYSOL from the $P2_1$ crystal structure (red line) produces a fit ($\chi = 1.362$) with visible deviations between the data, especially evident at low angles, suggesting that the crystal structure is more compact than that of pectocin M2 in solution.
B. Derivation of R_g from a Guinier analysis (red) of the scattering curve; residuals of the fit are in green.
C. Pair-distance distribution plot from experimental scattering data for pectocin M2 exhibiting two maxima which highlights the bimodal character of the molecule in solution. The D_{max} of the particle is 96 Å.
D. Porod-Debye and (E) normalized Kratky plots for pectocin M2 imply increased flexibility of the protein in solution (red). Pyocin M (black), a protein of similar molecular weight with a relatively rigid structure and strong inter-domain contacts is used as a control.

1995), from the pectocin M2 crystal structure shows there are obvious differences between the theoretical curve and experimental scattering data (Fig. 3A). In addition, the radius of gyration ($R_g = 27$ Å) obtained from Guinier analysis of the experimental scattering data is somewhat larger than that calculated from the pectocin M2 crystal structure ($R_g = 24$ Å) using SOMO (Rai *et al.*, 2005) (Fig. 3B). Consistent with this, the p(r) function, which describes the paired set of vectors between all the electrons within the protein, indicates a maximum particle size ($D_{max} = 96$ Å, Fig. 3C) that is much greater than the maximum dimension of the pectocin M2 crystal structure (77 Å, Fig. 1B). These data suggest that the pectocin M2 crystal structure is not wholly representative of the conformational ensemble present in solution and that this protein adopts an elongated conformation, implying inter-domain flexibility.

To test this idea further, we examined the Porod-Debye plot for pectocin M2, where scattering decay is examined as $I(q)q^4$ as a function of q^4. This analysis reports directly on particle flexibility and typically for compact globular particles an asymptotic plateau is reached for the low q part of the data. However, for pectocin M2 no discernible plateau was observed (Fig. 3D). For comparison, we also obtained scattering data for pyocin M which, as with colicin M and syringacin M, forms a compact structure and similarly analysed these data (Barreteau *et al.*, 2012b). In contrast to the curve obtained for pectocin M2, the Porod-Debye plot for pyocin M reached a plateau confirming its rigidity and compactness (Fig. 3D). In addition, the Kratky plot [$I(q)q^2$ versus q] for pectocin M2 normalized to the scattering intensity $I(0)$ and R_g, has two maxima with increasing scattering at higher angles. The Kratky plot reports directly on inter-domain flexibility and for pectocin M2 is consistent with a two-domain protein connected by a flexible linker (Fig. 3E). In comparison, there is a single maximum in the pyocin M Kratky plot, consistent with its single domain-like globular structure. Taken together these analyses indicate that pectocin M2 is flexible and adopts conformations distinct from that observed in the crystal structure.

Pectocin M2 can adopt a highly extended conformation and exists as two distinct subpopulations in solution

To determine if the SAXS data for pectocin M2 could be better described by an ensemble of conformations we first used discrete molecular dynamics (DMD) simulations

(Shirvanyants *et al.*, 2012) to explore the accessible conformational states of pectocin M2 and generated a random pool of 5000 possible conformationals using the crystal structure of pectocin M2. Next, we used a genetic algorithm implemented in the program GAJOE (Petoukhov *et al.*, 2012) to select for ensembles of these models that would better describe our SAXS data. Model selection was successful as judged by the close correlation of the theoretical scattering curve generated from the selected ensemble with the experimental SAXS data (Fig. 4A), indicating that our scattering data are best described by an ensemble of pectocin M2 conformers in solution. Interestingly, the selected ensembles show a bimodal distribution in comparison with the random pool of DMD-generated pectocin M2 models when the population frequency is plotted against R_g or D_{max} (Fig. 4B and C). Thus, in the population of selected conformations, we frequently find a compact conformation described by the first peak (with maxima at approximately 23 and 75 Å for R_g and D_{max} respectively) that approximates closely to the conformation found in the pectocin M2 $P2_1$ crystal structure for which R_g and D_{max} were calculated as 24 and 77 Å respectively. The second peak represents an ensemble of pectocin M2 conformers in an extended conformation with D_{max} values ranging up to 98 Å, which correlates closely with the experimentally determined value of D_{max} (96 Å). These analyses suggest that pectocin M2 can adopt both bent and elongated linear conformations in solution.

The bimodal distribution of the selected ensembles suggests discrete populations in solution, the more compact of which is similar to the conformation observed in the $P2_1$ crystals of pectocin M2. In an attempt to capture the more elongated conformation *in crystallo*, thus validating our solution scattering and modelling data, we repeated crystallization of pectocin M2. A custom re-crystallization screen was devised exploiting information from the initial trails. Crystals were obtained in a number of conditions from this screen and were tested for diffraction as well as space group and unit cell variation, which is indicative of novel packing. A form with the radically different space group of $P3_121$ was chosen for optimization, which yielded crystals diffracting to 1.86 Å. As an alternative domain arrangement to the $P2_1$ form was expected, data from this crystal form were again phased using anomalous data from the metal centres of the [2Fe-2S] cluster. During model building from these data it was immediately apparent that in this crystal form pectocin M2 did indeed adopt an elongated conformation (Fig. 5A). The calculated R_g and D_{max} for this structure were 28 and 97 Å respectively. These values correlate well with the extended population from the DMD simulation, suggesting that this structure is representative of the second elongated pool indentified by our modelling. Alignment of this elongated ($P3_121$) form and the original $P2_1$ form, based on their cytotoxic or ferredoxin

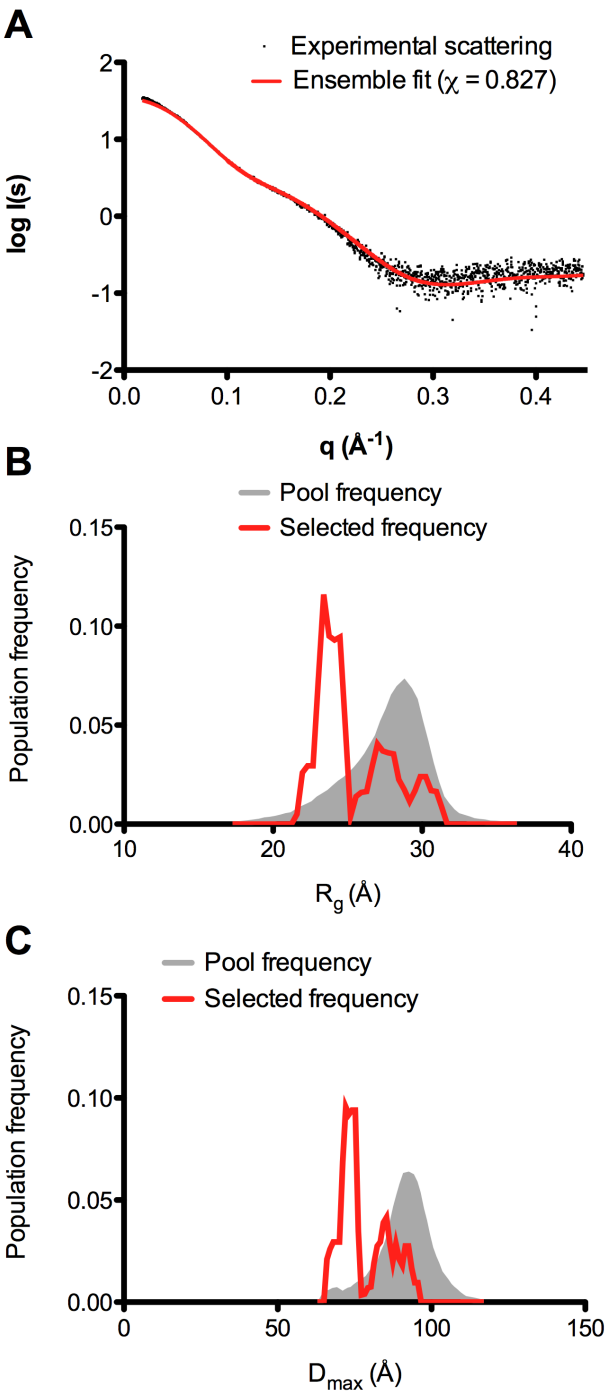


Fig. 4. Analysis of conformational heterogeneity of pectocin M2 reveals compact and extended ensembles in solution.
A. Overlay of scattering curves of pectocin M2 between experimental data and the best ensemble selected by GAJOE indicates improved fit to the scattering data ($\chi = 0.827$).
B and C. R_g (B) and D_{max} (C) distribution of solution ensembles selected by a genetic algorithm using GAJOE from a pool of 5000 random conformers of pectocin M2. Compact and elongated molecule were both selected implying that the protein is conformationally heterogeneous in solution allowing for significant inter-domain re-arrangements about the linker helix (residues 96–115).

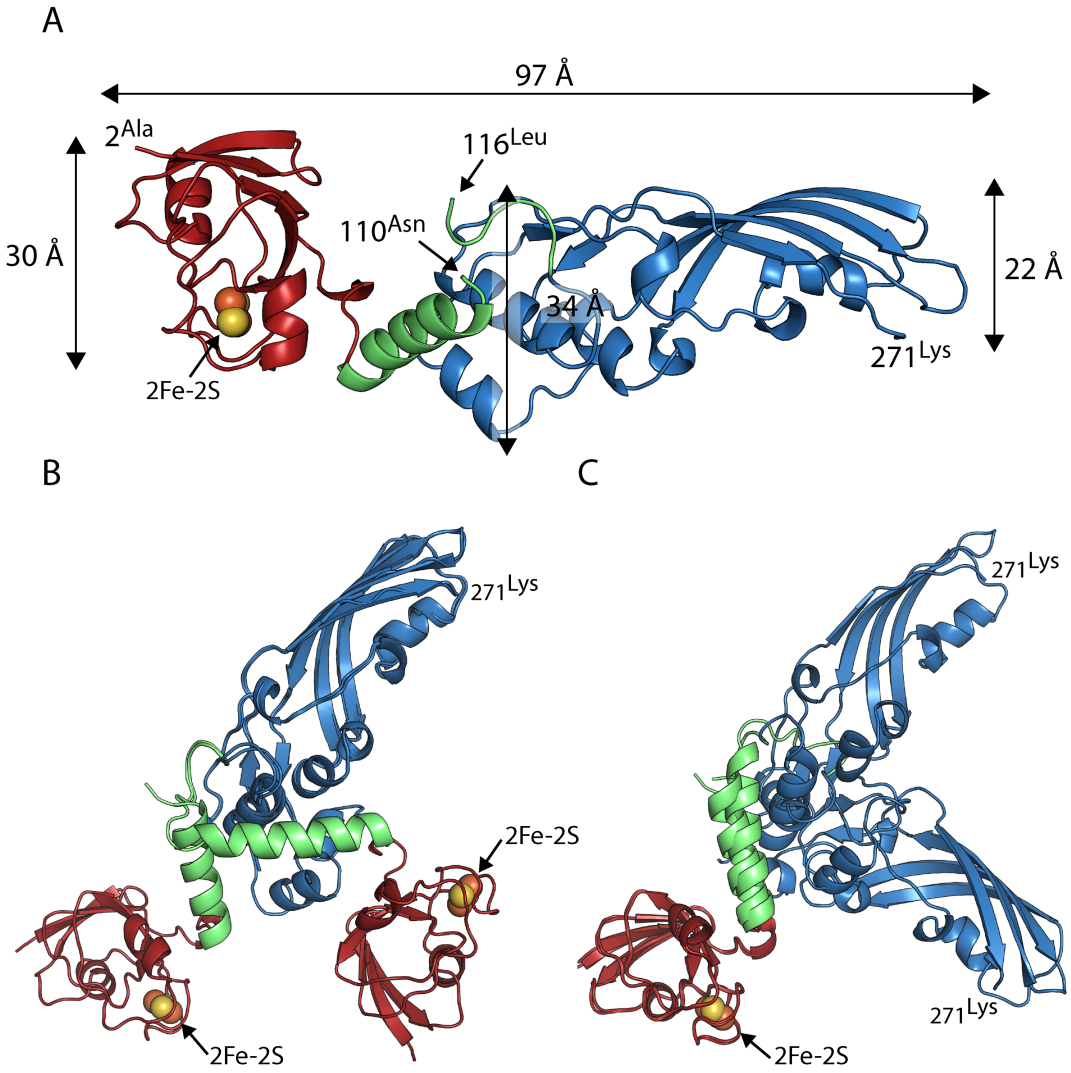


Fig. 5. Pectocin M2 $P3_121$ structure confirms extended conformation predicted by solution scattering.
A. Pectocin M2 in $P3_121$ crystal form is highly elongated, consistent with the extended conformation predicted from solution scattering and DMD simulations.
B. Alignment of the catalytic domains of the $P2_1$ and $P3_121$ crystal forms of pectocin M2, illustrating the difference in orientation between the ferredoxin and linker regions.
C. Alignment of the ferredoxin domains of the $P2_1$ and $P3_121$ crystal forms of pectocin M2, illustrating the difference in orientation of the catalytic domains.

domains, show a major difference in the relative orientations of these domains (Fig. 5B and C).

The ferredoxin domain is a generic module for the delivery of cytotoxic domains to the periplasm

In addition to pectocin M1 and M2, we previously identified a putative third member of the ferredoxin-containing bacteriocin family, designated pectocin P (Grinter *et al.*, 2012a). The open reading frame for pectocin P, identified in the genome of *Pectobacterium carotovorum* subsp. *carotovorum* WPP14, consists of an N-terminal ferredoxin domain, connected to a pesticin-like cytotoxic domain, which is analogous to T4 lysozyme. Similar to pectocins M1 and M2, there is no sequence N-terminal of the ferredoxin domain, so this bacteriocin also lacks an IUTD. To confirm that this open-reading frame encodes an active bacteriocin, we tested the cytotoxic activity of recombinantly expressed and purified pectocin P against diverse *Pectobacterium* isolates. For this test we utilized a solid growth inhibition assay conducted in parallel with pectocins M1 and M2 (Fyfe *et al.*, 1984). As with pectocins M1 and M2, limited inhibition of growth was observed under iron-replete conditions (LB agar). However, under iron-limiting

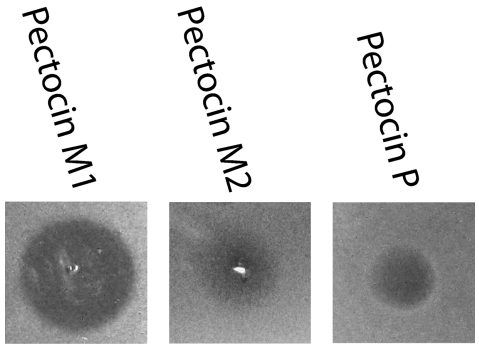


Fig. 6. Activity of pectocin P against *Pectobacterium atrosepticum* LMG 2386. Purified pectocins M1, M2 and P (2 mg ml^{-1}) were spotted onto a soft agar overlay seeded with *Pba* LMG 2386. Clear zones indicate inhibition of growth due to the cytotoxic effect of pectocins on cells.

conditions, inhibition of growth was observed for 17 of the 19 strains (Fig. 6, Table S1). The existence of pectocin P, an additional ferredoxin-containing bacteriocin with no N-terminal IUTD and a pesticin-like cytotoxic domain, provides strong supporting evidence that the ferredoxin domain acts as a generic module for the delivery of cytotoxic domains to the periplasm in *Pectobacterium* spp. The cytotoxic domains of both M-class bacteriocins and pesticin have been studied extensively and there is no indication that they possess any intrinsic capacity to cross the OM (Barreteau *et al*., 2010; Helbig and Braun, 2011; Lukacik *et al*., 2012; Patzer *et al*., 2012), indicating that ferredoxin uptake represents an unprecedented example of receptor-mediated protein uptake for nutrient acquisition in bacteria.

Discussion

In this work we present the structure and solution properties of the atypical bacteriocin pectocin M2, which consists of a fusion between a colicin M-like cytotoxic domain and a plant-derived ferredoxin domain. We have previously demonstrated that *Pectobacterium* spp. are able to acquire iron directly from plant ferredoxin under iron-limiting conditions through a receptor mediated process and that the bacteriocins pectocin M1 and M2 parasitize this system for cell entry through presentation of a ferredoxin domain (Grinter *et al*., 2013). In this study we provide an insight into how this uptake occurs by showing that pectocin M2 has an unprecedented structure among bacteriocins in that it lacks an IUTD. Additionally, we definitively show that pectocin M2 is highly flexible in solution fluctuating between compact and extended conformations.

All iron-uptake systems identified to date in Gram-negative bacteria, either siderophore based or targeting a protein substrate, utilize a TonB-dependent receptor to transport iron across the outer membrane (Faraldo-Gomez and Sansom, 2003). Likewise, binding and parasitization of these receptors for cell entry is a characteristic trait of colicin M, and other characterized colicins and pyocins, including the E-type colicins (E1-E9), colicins A, B, D, IA and IB and pyocins S1-S5 (Loftus *et al*., 2006; Buchanan *et al*., 2007; Cascales *et al*., 2007; Denayer *et al*., 2007; Devanathan and Postle, 2007; Elfarash *et al*., 2012). As such, while it is yet to be confirmed that a TonB-dependent receptor is responsible for mediating ferredoxin iron and pectocin uptake, this class of protein is by far the most likely candidate.

In bacteriocins the IUTD normally functions to deliver an epitope to the periplasm, which mediates binding to the Tol or Ton complexes. This direct interaction occurs between the colicins and TolB for group A colicins and TonB for group B colicins and is essential for uptake of these bacteriocins. In addition to their subversion for bacteriocin import, Tol and Ton complexes have a general physiological role in the bacterial cell. The Ton complex provides the energy required for the import of iron containing siderophores and related substrates through TonB-dependent receptors, by interaction with the receptor plug domain subsequent to binding of the substrate on the outer surface of the receptor (Noinaj *et al*., 2010). Given the universal role of TonB-dependent receptors in iron transport across the outer membrane and their parasitization by colicin-like bacteriocins, it is reasonable to speculate that the ferredoxin/pectocin receptor is a member of this class of protein.

Since pectocin M1 and M2 lack an IUTD they are unable make direct contact with the Tol or Ton complexes in the periplasm and thus are unable to directly utilize the pmf for cell entry. However, the fact that these proteins parasitize a system for which the receptor binding and translocation domains are structurally analogous to the substrate provides an intuitive solution to this problem. A number of TonB-dependent receptors have been identified, which obtain iron from host proteins during infection. In all of these systems the iron or iron containing compound is liberated from the protein on the cell surface and transported into the cell, potentially because all of the proteins identified are too large pass through the lumen of their receptor (Wandersman and Stojiljkovic, 2000; Faraldo-Gomez and Sansom, 2003; Noinaj *et al*., 2012). Plant ferredoxin, however, is a small globular protein, which is in fact comparable in dimensions to the plug domain that ordinarily occludes the pore of a TonB-dependent receptor (Fig. 7A and C). This creates the possibility that the ferredoxin is imported intact into the periplasm. If this were the case it could readily explain the pectocins lack of an IUTD as the energy required for cell entry would still be provided by the Ton complex, but

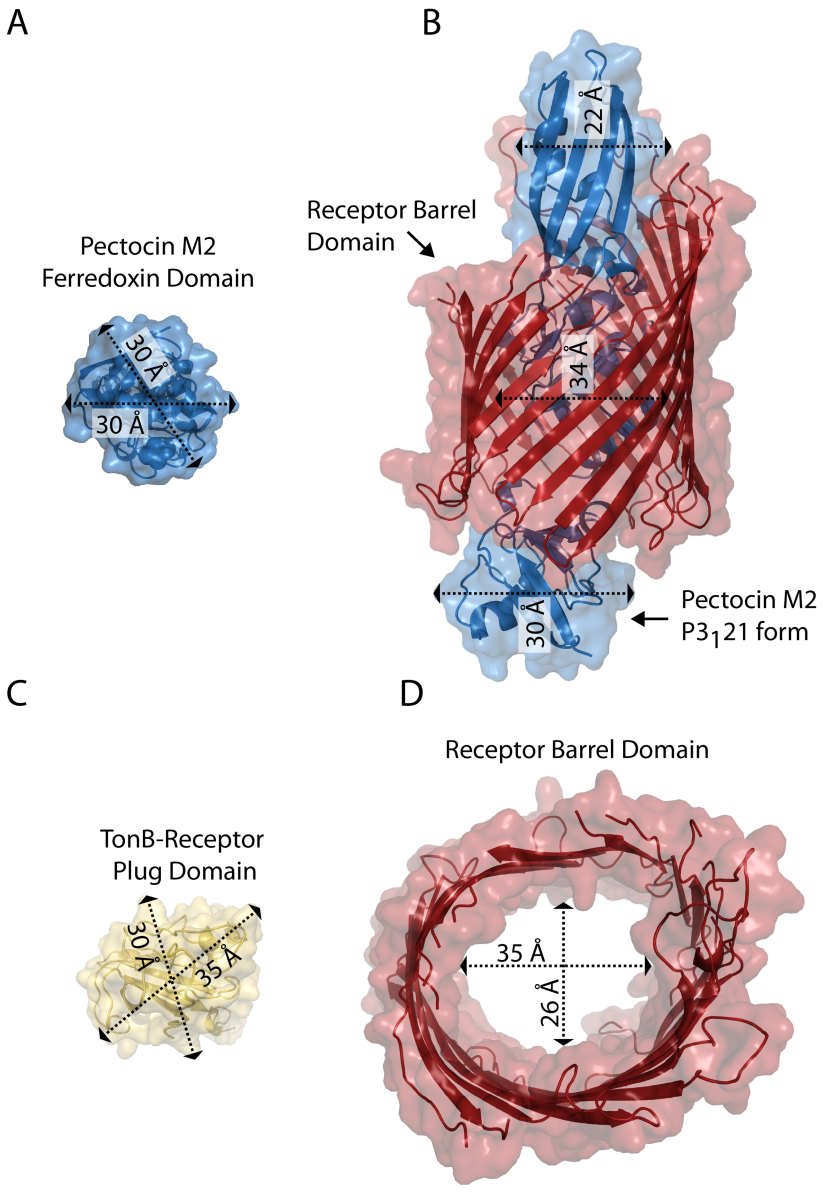


Fig. 7. The extended conformation of pectocin M2 has dimensions compatible with passage through the lumen of a TonB-dependent receptor.
A. The extended conformation of pectocin M2 (blue cartoon and surface), fitted into the pore of the barrel of HasR from *Serratia marcescens* PDB ID = 3CLS, illustrating that the extended conformation of pectocin M2 is conducive to transport through the lumen of a TonB-dependent receptor without unfolding.
B. Dimensions of the ferredoxin domain of pectocin M2.
C. The width of the plug domain, which ordinarily blocks the lumen of the receptor barrel, is similar to that of the elongated conformation of pectocin M2.
D. Top-down view of the HasR barrel showing the internal dimensions of the barrel domain.

transduced through receptor plug domain, as with ordinary substrate importation.

As the elongated conformation of pectocin M2 has comparable dimensions along its length (Fig. 5A) we manually docked this model into the barrel domain of the hemophore receptor HasR, a TonB-dependent receptor shown to be responsible for importation of the relatively bulky substrate, haem (Fig. 7B and D). This docking illustrates that pectocin M2 in its elongated conformation could traverse the lumen of such a receptor to gain entry to the cell, with the flexibility of pectocin M2 observed in solution allowing the protein to adapt to the shape of the lumen of its transporter during importation. In the case of the transport of iron-siderophores, it is generally thought that there are two possibilities with respect to the role of TonB in stimulating substrate transport. Either TonB induces a rearrangement of the plug domain within the barrel that is sufficient to allow passage of the substrate or it directly pulls the plug domain completely from the barrel (Usher *et al.*, 2001; Udho *et al.*, 2009; 2012). However, for an intact protein such as ferredoxin to be translocated directly through the lumen of a TonB-dependent receptor, it would be necessary for the plug domain to completely exit the barrel during substrate transport and it has not as yet been directly demonstrated that this occurs in TonB-dependent receptors. The identification of this receptor and the testing of this importation hypothesis represents an intriguing question for future work.

Experimental procedures

Expression and purification of ferredoxins and pectocins

A list of bacterial strains and plasmids used in this work is provided in Table S1. The open reading frame for pectocin P minus stop codon was synthesized by DNA 2.0 and ligated into the expression vector pJ404 (T5 promoter, C-terminal His_6-tag). The resulting vector was designated pJPP1. Vectors for expression of pectocins M1 and M2 were as previously described (Grinter *et al*., 2012a). All proteins were expressed and purified as previously described (Grinter *et al*., 2012a). Proteins were concentrated to 5–15 mg ml^{-1} using a centrifugal concentrator and stored at −80°C in this buffer until required. For pectocins M1 and P, 5% glycerol was added to all buffers as it was found to enhance protein stability.

Cytotoxicity assays

The cytotoxicity of purified pectocins was tested using the soft agar overlay method (Fyfe *et al*., 1984). 200 μl of mid-log phase culture of the test strain was added to 6 ml of 0.6% agar melted and cooled to 42°C. The molten agar was then overlaid onto LB medium with or without 100–400 μM 2,2′-biyridine. Purified pectocins (2 mg ml^{-1}) were spotted directly onto the surface of the overlay, once solidified. Plates were incubated at 28°C for 16 h, and monitored for zones of growth inhibition.

SAXS data collection and analysis

SAXS data were collected on the X33 beamline at the Deutsches Elektronen Synchrotron (DESY, Hamburg, Germany). Pectocin M2 and pyocin M concentrations between 0.4 and 4.0 mg ml^{-1} were used. Average buffer scattering was subtracted from the sample scattering. The first 200 points (low angle data) of the scattering curve obtained for 1 mg ml^{-1} protein were merged with the rest of the high angle data from the 4 mg ml^{-1} sample to avoid the influence on the data of any inter-particle interference. All data processing was performed using PRIMUS (Konarev *et al*., 2003). Porod-Debye [$I(q)q^4$ versus q^4] and normalized Kratky [$I(q)q^2$ versus q] plots were used to assess particle flexibility as described in the *Results* section (Durand *et al*., 2010; Rambo and Tainer, 2011). The distance distribution function, p(r), was obtained by indirect Fourier transform of the scattering intensity using GNOM (Svergun, 1992). A Guinier plot [ln $I(s)$ versus q^2] was used to determine the radius of gyration, R_g, of pectocin M2 and pyocin M. US-SOMO (Rai *et al*., 2005) was used to determine hydrodynamic parameters based on the crystal structures of pectocin M2. CRYSOL (Svergun *et al*., 1995) was used to compute theoretical scattering curves from high-resolution X-ray structures.

DMD and EOM simulations

Discrete molecular dynamics simulations of the pectocin M2 linker region (residues 96–116) were undertaken with US-SOMO (Brookes *et al*., 2010; Shirvanyants *et al*., 2012) in order to explore pectocin M2 conformational space. The Andersen thermostat temperature (T) was set to 0.5 kcal mol^{-1} K^{-1} to allow for sufficient sampling of conformational dynamics around the native state without melting the structure of the linker. The run time and pdb output step were adjusted in order to generate 5000 models. Next, the pool of 5000 'random' pectocin M2 conformers generated was refined by a genetic algorithm implemented in the program GAJOE as part of the ensemble optimization method (Bernadó *et al*., 2007; Petoukhov *et al*., 2012).

Pectocin M2 crystallization and diffraction data collection

Initial crystallization trials were performed at the high throughput crystallization facility of the University of Zurich using the vapour diffusion method (reservoir volume of 50 μl, drop size: 100 nl protein, 100 nl reservoir solution) with pectocin M2 at a final concentration of 15 mg ml^{-1}. Pectocin M2 formed crystals or spherulites in a number of conditions containing ammonium sulphate and PEG 3350. Crystals were extracted from one of these conditions [15% PEG 3350, 0.2 M ammonium sulphate, 3% 2-methyl-2,4-pentanediol (MPD), 0.1 M bis-tris pH 6.5], cryoprotected by increasing PEG 3350 to 30% and data were collected at 100 K to 2.3 Å in the space group $P2_1$, at the SLS (Zurich). Re-crystallization screening of pectocin M2 was performed, using a custom screen with variations in the concentration/ratio of precipitants from the original condition (ammonium sulphate and PEG 3350), pH and additives. Clusters of large rod-shaped crystals formed at high ammonium sulphate concentrations. This was optimized giving a final condition of 1.8 M ammonium sulphate, 3% MPD, 0.1 M MES, pH 6.5. These crystals were manually separated and cryoprotected with 15–20% glycerol. Data were collected at 100 K on beamlines I02 and I03 at the Diamond Light Source (Oxfordshire, UK). Automatic data-processing was performed with Xia2 within the EDNA package (Incardona *et al*., 2009). Datasets for experimental phasing using the iron–sulphur cluster of pectocin M2 were collected at the iron K-edge (1.7433 Å) and high-resolution data were collected at 0.9796 Å. Data collection statistics from both crystal forms are reported in Table 1.

Pectocin M2 structure solution and refinement

Phases for the $P2_1$ and $P3_121$ datasets were obtained from the anomalously scattering substructure from the pectocin–ferredoxin domain iron–sulphur cluster, determined for the iron-edge dataset using the Hybrid Substructure Search from the Phenix package (Adams *et al*., 2010). Four positions were located per ASU corresponding to two iron–sulphur clusters (correlation coefficient = 0.5) from two pectocin M2 molecules. These positions were then utilized by Phaser-EP (McCoy *et al*., 2007; Adams *et al*., 2010; Winn *et al*., 2011) phases were improved using RESOLVE density modification from the Phenix package, and the initial model was built and refined using Phenix Autobuild (Adams *et al*., 2010). The model was then built and refined manually using Coot 0.7 and Refmac5

Table 1. Crystallographic data collection and refinement statistics.

		High resolution dataset	Iron edge dataset
Data collection[a]			
Space group	$P2_1$	$P3_121$	$P3_121$
Cell dimensions			
a, *b*, *c* (Å)	44.65, 116.75, 60.78	117.45, 117.45, 128.45	117.26, 117.26, 128.53
α, β, γ (°)	90, 94.96, 90	90, 90, 120	90, 90, 120
Resolution (Å)	50.00–2.30 (2.44–2.30)	64.22–1.86 (1.91–1.86)	43.31–2.01 (2.06–2.01)
R_{merge} (%)	5.0 (59.9)	3.4 (68.3)	3.8 (69.3)
R_{pim} (%)[b]	–	0.9 (15.7)	2.0 (31.0)
Mean *I*/sigma(*I*)	12.41 (1.85)	47.1 (5.5)	28.9 (3.1)
Completeness (%)	96.6 (95.7)	100.0 (99.9)	99.4 (98.7)
Redundancy	3.2 (2.8)	20 (20.8)	8 (6.6)
Refinement			
Resolution (Å)	50.00–2.30 (2.44–2.30)	64.22–1.86 (1.91–1.86)	
No. of reflections	52652 (8438)	86092 (6304)	
R_{work}/R_{free} (%)	21.0/27.2	16.8/19.1	
No. of atoms			
Protein	4093	4175	
Ligand/ion	34	388	
Water	49	438	
B factors			
Protein	69.1	45.5	
2Fe-2S	74.6	35.4	
$SO4^{2-}/Cl^-$	95	68.2	
Glycerol/MPD	–	70.7	
Water	42.4	56.0	
Root mean square deviations			
Bond lengths (Å)	0.016	0.024	
Bond angles (°)	1.757	2.55	
PDB identifier	4N59	4N58	

a. Values in parentheses refer to the highest resolution shell.
b. $R_{pim} = \Sigma_{hkl}[1/(N-1)^{1/2}\Sigma_i|I_i(hkl) - <I(hkl)>|/\Sigma_{hkl}\Sigma_i I_i(hkl)$

(McCoy *et al.*, 2007; Emsley *et al.*, 2010; Murshudov *et al.*, 2011). Validation of refined structures was performed using the Molprobity web server and Procheck from CCP4i (Laskowski *et al.*, 1993; Chen *et al.*, 2010). Refinement statistics for both datasets are reported in Table 1.

Lipid II cleavage assay

Lipid II hydrolysis assays were performed and visualized as previously described by (Grinter *et al.*, 2012b), with non-radiolabelled lipid II substrate obtained from the UK Bacterial Cell Wall Biosynthesis Network (Lloyd *et al.*, 2008; Clarke *et al.*, 2009). A band corresponding to lipid II was observed with an R_f of 0.7 as reported previously (Barreteau *et al.*, 2009).

Acknowledgements

We thank the Diamond Light Source for access to beamlines I02 and I03 (proposal numbers MX6638 and MX8659), We acknowledge funding from P-CUBE for initial crystallization trials performed at HT-X Zurich and were are grateful to Beat Blattmann for technical assistance. R.G. is supported by a Kelvin-Smith Scholarships from the University of Glasgow. I.J. and L.M. are supported by studentships from the Wellcome Trust, award numbers 093592/Z/10/Z and 093597/Z/10/Z respectively.

References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66:** 213–221.

Barreteau, H., Bouhss, A., Fourgeaud, M., Mainardi, J.-L., Touzé, T., Gérard, F., *et al.* (2009) Human- and plant-pathogenic pseudomonas species produce bacteriocins exhibiting colicin M-like hydrolase activity towards peptidoglycan precursors. *J Bacteriol* **191:** 3657–3664.

Barreteau, H., Bouhss, A., Gérard, F., Duché, D., Boussaid, B., Blanot, D., *et al.* (2010) Deciphering the catalytic domain of colicin M, a peptidoglycan lipid II-degrading enzyme. *J Biol Chem* **285:** 12378–12389.

Barreteau, H., Ghachi, M.E., Barnéoud-Arnoulet, A., Sacco, E., Touzé, T., Duché, D., *et al.* (2012a) Characterization of colicin M and its orthologs targeting bacterial cell wall peptidoglycan biosynthesis. *Microb Drug Resist* **18:** 222–229.

Barreteau, H., Tiouajni, M., Graille, M., Josseaume, N., Bouhss, A., Patin, D., *et al.* (2012b) Functional and structural characterization of PaeM, a colicin M-like bacteriocin produced by *Pseudomonas aeruginosa*. *J Biol Chem* **287:** 37395–37405.

Bernadó, P., Mylonas, E., Petoukhov, M.V., Blackledge, M., and Svergun, D.I. (2007) Structural characterization of flex-

ible proteins using small-angle X-ray scattering. *J Am Chem Soc* **129:** 5656–5664.
Brookes, E., Demeler, B., Rosano, C., and Rocco, M. (2010) The implementation of SOMO (SOlution MOdeller) in the UltraScan analytical ultracentrifugation data analysis suite: enhanced capabilities allow the reliable hydrodynamic modeling of virtually any kind of biomacromolecule. *Eur Biophys J* **39:** 423–435.
Buchanan, S.K., Lukacik, P., Grizot, S., Ghirlando, R., Ali, M.M.U., Barnard, T.J., *et al*. (2007) Structure of colicin I receptor bound to the R-domain of colicin Ia: implications for protein import. *EMBO J* **26:** 2594–2604.
Cascales, E., Buchanan, S.K., Duché, D., Kleanthous, C., Lloubès, R., Postle, K., *et al*. (2007) Colicin biology. *Microbiol Mol Biol Rev* **71:** 158–229.
Chen, V.B., Arendall, W.B., III, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., *et al*. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66:** 12–21.
Clarke, T.B., Kawai, F., Park, S.-Y., Tame, J.R.H., Dowson, C.G., and Roper, D.I. (2009) Mutational analysis of the substrate specificity of *Escherichia coli* penicillin binding protein 4. *Biochemistry* **48:** 2675–2683.
Denayer, S., Matthijs, S., and Cornelis, P. (2007) Pyocin S2 (Sa) kills *Pseudomonas aeruginosa* strains via the FpvA Type I ferripyoverdine receptor. *J Bacteriol* **189:** 7663–7668.
Devanathan, S., and Postle, K. (2007) Studies on colicin B translocation: FepA is gated by TonB. *Mol Microbiol* **65:** 441–453.
Durand, D., Vivès, C., Cannella, D., Pérez, J., Pebay-Peyroula, E., Vachette, P., and Fieschi, F. (2010) NADPH oxidase activator p67phox behaves in solution as a multi-domain protein with semi-flexible linkers. *J Struct Biol* **169:** 45–53.
Elfarash, A., Wei, Q., and Cornelis, P. (2012) The soluble pyocins S2 and S4 from *Pseudomonas aeruginosa* bind to the same FpvAI receptor. *MicrobiologyOpen* **1:** 268–275.
Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66:** 486–501.
Faraldo-Gomez, J.D., and Sansom, M.S.P. (2003) Acquisition of siderophores in Gram-negative bacteria. *Nat Rev Mol Cell Biol* **4:** 105–116.
Fyfe, J.A., Harris, G., and Govan, J.R. (1984) Revised pyocin typing method for *Pseudomonas aeruginosa*. *J Clin Microbiol* **20:** 47–50.
Grinter, R., Milner, J., and Walker, D. (2012a) Ferredoxin containing bacteriocins suggest a novel mechanism of iron uptake in *Pectobacterium* spp. *PLoS ONE* **7:** e33033.
Grinter, R., Roszak, A.W., Cogdell, R.J., Milner, J.J., and Walker, D. (2012b) The crystal structure of the lipid II-degrading bacteriocin syringacin M suggests unexpected evolutionary relationships between colicin M-like bacteriocins. *J Biol Chem* **287:** 38876–38888.
Grinter, R., Milner, J., and Walker, D. (2013) Beware of proteins bearing gifts: protein antibiotics that use iron as a Trojan horse. *FEMS Microbiol Lett* **338:** 1–9.
Helbig, S., and Braun, V. (2011) Mapping functional domains of colicin M. *J Bacteriol* **193:** 815–821.
Housden, N.G., Wojdyla, J.A., Korczynska, J., Grishkovskaya, I., Kirkpatrick, N., Brzozowski, A.M., and Kleanthous, C. (2010) Directed epitope delivery across the *Escherichia coli* outer membrane through the porin OmpF. *Proc Natl Acad Sci USA* **107:** 21412–21417.
Housden, N.G., Hopper, J.T.S., Lukoyanova, N., Rodriguez-Larrea, D., Wojdyla, J.A., Klein, A., *et al*. (2013) Intrinsically disordered protein threads through the bacterial outer-membrane porin OmpF. *Science* **340:** 1570–1574.
Incardona, M.-F., Bourenkov, G.P., Levik, K., Pieritz, R.A., Popov, A.N., and Svensson, O. (2009) EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. *J Synchrotron Radiat* **16:** 872–879.
Jakes, K.S., and Finkelstein, A. (2010) The colicin Ia receptor, Cir, is also the translocator for colicin Ia. *Mol Microbiol* **75:** 567–578.
Kleanthous, C. (2010) Swimming against the tide: progress and challenges in our understanding of colicin translocation. *Nat Rev Microbiol* **8:** 843–848.
Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* **36:** 1277–1282.
Laskowski, R.A., Macarthur, M.W., Moss, D.S., and Thornton, J.M. (1993) PROCHECK – a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26:** 283–291.
Lloyd, A.J., Gilbey, A.M., Blewett, A.M., De Pascale, G., El Zoeiby, A., Levesque, R.C., *et al*. (2008) Characterization of tRNA-dependent peptide bond formation by MurM in the synthesis of *Streptococcus pneumoniae* peptidoglycan. *J Biol Chem* **283:** 6402–6417.
Loftus, S.R., Walker, D., Maté, M.J., Bonsor, D.A., James, R., Moore, G.R., and Kleanthous, C. (2006) Competitive recruitment of the periplasmic translocation portal TolB by a natively disordered domain of colicin E9. *Proc Natl Acad Sci USA* **103:** 12353–12358.
Lukacik, P., Barnard, T.J., Keller, P.W., Chaturvedi, K.S., Seddiki, N., Fairman, J.W., *et al*. (2012) Structural engineering of a phage lysin that targets Gram-negative pathogens. *Proc Natl Acad Sci USA* **109:** 9857–9862.
McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007) Phaser crystallographic software. *J Appl Crystallogr* **40:** 658–674.
Michel-Briand, Y., and Baysse, C. (2002) The pyocins of *Pseudomonas aeruginosa*. *Biochimie* **84:** 499–510.
Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., *et al*. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67:** 355–367.
Noinaj, N., Guillier, M., Barnard, T.J., and Buchanan, S.K. (2010) TonB-dependent transporters: regulation, structure, and function. *Annu Rev Microbiol* **64:** 43–60.
Noinaj, N., Easley, N.C., Oke, M., Mizuno, N., Gumbart, J., Boura, E., *et al*. (2012) Structural basis for iron piracy by pathogenic *Neisseria*. *Nature* **483:** 53–58.
Patzer, S.I., Albrecht, R., Braun, V., and Zeth, K. (2012) Structure and mechanistic studies of pesticin, a bacterial homolog of phage lysozymes. *J Biol Chem* **287:** 23381–23396.
Petoukhov, M.V., Franke, D., Shkumatov, A.V., Tria, G., Kikhney, A.G., Gajda, M., *et al*. (2012) New developments

in the ATSAS program package for small-angle scattering data analysis. *J Appl Crystallogr* **45:** 342–350.
Rai, N., Nöllmann, M., Spotorno, B., Tassara, G., Byron, O., and Rocco, M. (2005) SOMO (SOlution MOdeler): differences between X-ray- and NMR-derived bead models suggest a role for side chain flexibility in protein hydrodynamics. *Structure* **13:** 723–734.
Rambo, R.P., and Tainer, J.A. (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95:** 559–571.
Riley, M.A., Pinou, T., Wertz, J.E., Tan, Y., and Valletta, C.M. (2001) Molecular characterization of the klebicin B plasmid of *Klebsiella pneumoniae*. *Plasmid* **45:** 209–221.
Shirvanyants, D., Ding, F., Tsao, D., Ramachandran, S., and Dokholyan, N.V. (2012) Discrete molecular dynamics: an efficient and versatile simulation method for fine protein characterization. *J Phys Chem B* **116:** 8375–8382.
Svergun, D. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* **25:** 495–503.
Svergun, D., Barberato, C., and Koch, M.H.J. (1995) CRYSOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* **28:** 768–773.
Udho, E., Jakes, K.S., Buchanan, S.K., James, K.J., Jiang, X., Klebba, P.E., and Finkelstein, A. (2009) Reconstitution of bacterial outer membrane TonB-dependent transporters in planar lipid bilayer membranes. *Proc Natl Acad Sci USA* **106:** 21990–21995.
Udho, E., Jakes, K.S., and Finkelstein, A. (2012) TonB-dependent transporter FhuA in planar lipid bilayers: partial exit of its plug from the barrel. *Biochemistry* **51:** 6753–6759.
Usher, K.C., Özkan, E., Gardner, K.H., and Deisenhofer, J. (2001) The plug domain of FepA, a TonB-dependent transport protein from *Escherichia coli*, binds its siderophore in the absence of the transmembrane barrel domain. *Proc Natl Acad Sci USA* **98:** 10676–10681.
Wandersman, C., and Stojiljkovic, I. (2000) Bacterial heme sources: the role of heme, hemoprotein receptors and hemophores. *Curr Opin Microbiol* **3:** 215–220.
Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67:** 235–242.
Zeth, K., Römer, C., Patzer, S.I., and Braun, V. (2008) Crystal structure of colicin M, a novel phosphatase specifically imported by *Escherichia coli*. *J Biol Chem* **283:** 25324–25331.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Acta Crystallographica Section F
Structural Biology Communications
ISSN 2053-230X

Recombinant expression, purification, crystallization and preliminary X-ray diffraction analysis of the C-terminal DUF490$_{963–1138}$ domain of TamB from *Escherichia coli*

Inokentijs Josts,[a] Rhys Grinter,[a] Sharon M. Kelly,[b] Khedidja Mosbahi,[a] Aleksander Roszak,[b,c] Richard Cogdell,[b] Brian O. Smith,[b] Olwyn Byron[d] and Daniel Walker[a]*

[a]Institute of Infection, Immunity and Inflammation, College of Medical, Veterinary and Life Sciences, University of Glasgow, Scotland, [b]Institute of Molecular Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Scotland, [c]WestCHEM, School of Chemistry, College of Science and Engineering, University of Glasgow, Scotland, and [d]School of Life Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Scotland

Correspondence e-mail: daniel.walker@glasgow.ac.uk

Received 4 June 2014
Accepted 29 July 2014

TamB is a recently described inner membrane protein that, together with its partner protein TamA, is required for the efficient secretion of a subset of autotransporter proteins in Gram-negative bacteria. In this study, the C-terminal DUF490$_{963–1138}$ domain of TamB was overexpressed in *Escherichia coli* K-12, purified and crystallized using the sitting-drop vapour-diffusion method. The crystals belonged to the primitive trigonal space group $P3_121$, with unit-cell parameters $a = b = 57.34$, $c = 220.74$ Å, and diffracted to 2.1 Å resolution. Preliminary secondary-structure and X-ray diffraction analyses are reported. Two molecules are predicted to be present in the asymmetric unit. Experimental phasing using selenomethionine-labelled protein will be undertaken in the future.

1. Introduction

Proteobacterial genomes are littered with proteins devoid of functional annotation. A function for TamB (previously YtfN) and its partner protein TamA (previously YtfM) has recently been elucidated (Selkrig *et al.*, 2012). Both proteins constitute the translocation and assembly module (TAM), a cell envelope-spanning protein complex found in most proteobacteria that contributes to the efficient secretion of autotransporter (AT) proteins (Selkrig *et al.*, 2012). ATs are secreted outer membrane proteins that possess a C-terminal outer membrane (OM)-bound 12-stranded β-barrel domain and an N-terminal passenger domain consisting of an extended β-helical structure (Nishimura *et al.*, 2010). The insertion of the β-domain into the OM is dependent on the β-barrel assembly machinery (BAM) complex, whereas the secretion of the passenger domain seemingly relies on the TAM complex for efficient OM translocation (Selkrig *et al.*, 2012; Sauri *et al.*, 2009). TamB is a large (137 kDa) multi-domain inner membrane protein for which little structural information is available. Amino-acid sequence analysis of TamB reveals the presence of a C-terminal conserved domain of unknown function DUF490 (Fig. 1*a*), which has previously been shown to associate with its partner protein TamA (Selkrig *et al.*, 2012). Bioinformatic analyses of the DUF490 domain reveal the widespread occurrence of this domain, and some examples of other proteins with this domain, as well as sequence alignments, are presented in Supplementary Fig. S1[1]. Structural characterization of the DUF490 domain is therefore of interest since it might aid in shedding light on the biological functions of TamB and the numerous other proteins of unknown function that contain this domain. Here, we report the cloning, expression, purification, crystallization and collection of X-ray crystallographic data for DUF490$_{963–1138}$ from the *Escherichia coli* K-12 TamB protein.

2. Materials and methods

2.1. Macromolecule production

DUF490$_{963–1138}$ was cloned from a TamB$_{DUF490}$ construct previously produced from *E. coli* strain K-12 (Selkrig *et al.*, 2012). The

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: WD5236).

amplified PCR product was digested with *Nde*I and *Xho*I and ligated into pET-21a to produce an expression plasmid which encodes DUF490$_{963–1138}$ with a C-terminal His$_6$ tag.

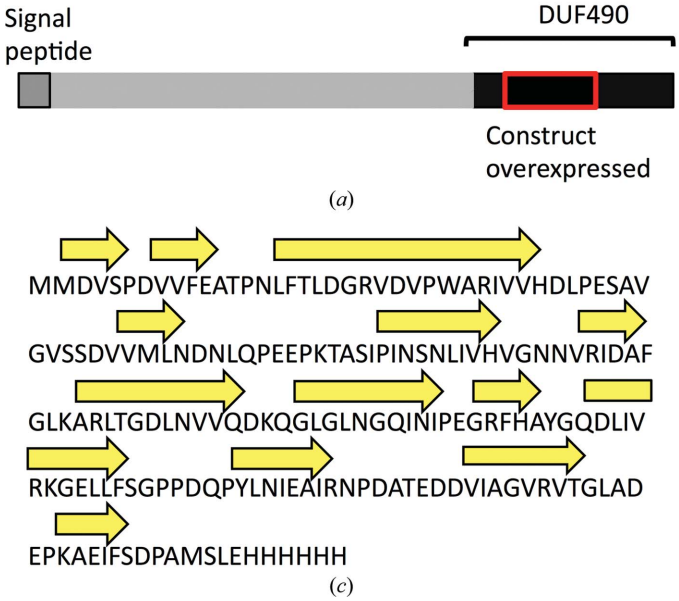
E. coli BL21 (DE3) cells transformed with the DUF490$_{963–1138}$ expression plasmid were grown in LB medium supplemented with 3%(*v*/*v*) glycerol and 100 mg ml^{-1} ampicillin at 37°C to an OD$_{600}$ of 0.4–0.6; isopropyl β-D-1-thiogalactopyranoside (IPTG) was then added to a final concentration of 0.5 m*M* to induce protein overexpression and the cells were grown for a further 15 h at 25°C. The cell pellet was collected by centrifugation at 4400*g* and resuspended in buffer *A* [20 m*M* Tris–HCl, 10 m*M* imidazole, 0.5 *M* NaCl, 5%(*v*/*v*) glycerol, 0.05% LDAO pH 7.5] supplemented with cOmplete protease-inhibitor cocktail (Roche) plus lysozyme (2 mg ml^{-1}) and lysed by sonication. Cell debris were cleared by additional centrifugation at 46 000*g* and the supernatant was passed through a nickel-charged HisTrap HP column (GE Healthcare). The bound fractions were collected after elution with buffer *B* [20 m*M* Tris–HCl, 350 m*M* imidazole, 0.5 *M* NaCl, 5%(*v*/*v*) glycerol, 0.05% LDAO pH 7.5]. Fractions containing the protein of interest were pooled, dialysed in 50 m*M* Tris, 200 m*M* NaCl, 3%(*v*/*v*) glycerol, 0.05% LDAO pH 7.5 and run on a Superdex S200 gel-filtration column equilibrated with the same buffer. After SEC, DUF490$_{963–1138}$ was dialysed into the same buffer omitting the 3% glycerol and stored at −80°C. Purified DUF490$_{963–1138}$ was analysed by far-UV circular-dichroism (CD) spectroscopy. CD measurements were obtained using a protein concentration of 25 μ*M* in a 0.02 cm quartz cuvette with a Jasco J-810 spectropolarimeter (Jasco UK Ltd). Secondary-structure estimates were obtained using the *CONTIN* procedure which was available from the *DichroWeb* server (Provencher & Glöckner, 1981; Whitmore & Wallace, 2008). Bioinformatic analysis of DUF490 was carried out using the *Conserved Domain Architecture Retrieval Tool* (*CDART*; Geer *et al.*, 2002). Macromolecule production is summarized in Table 1.

2.2. Crystallization

For crystallization, DUF490$_{963–1138}$ was concentrated to 15 mg ml^{-1} using Vivaspin ultracentrifugation spin columns (MWCO 4000–6000) and filtered prior to dispensing into crystallization trays. Initial crystallization screens were set up in a 96-well MRC sitting-drop vapour-diffusion format (60 µl reservoir solution, 0.5 µl protein + 0.5 µl reservoir) using a Cartesian Honeybee 8+1 dispensing robot. Crystals appeared after several days in 0.1 *M* HEPES, 15%(*v*/*v*) PEG 400, 0.2 *M* $CaCl_2$ pH 7.0 at 289 K. After optimization, diffraction-quality crystals were obtained in 0.1 *M* HEPES, 25%(*v*/*v*) PEG 400, 0.2 *M* $CaCl_2$ pH 8.0 using 8 mg ml^{-1} protein and a protein:reservoir ratio of 1:1 at 289 K. The crystals were cryoprotected using 0.1 *M* HEPES, 30%(*v*/*v*) PEG 400, 0.2 *M* $CaCl_2$ pH 8.0 by transferring the crystals using a LithoLoop into the cryosolution for 3 s and flash-cooling in a nitrogen-gas stream at 110 K. Crystallization is summarized in Table 2.

2.3. Data collection and processing

X-ray diffraction data were collected on beamline I03 at the Diamond Light Source (DLS) synchrotron, Harwell, England. Data were collected using a PILATUS3 6M detector with an oscillation angle of 0.15° and 0.08 s exposure time. A total of 3600 frames were collected and indexed using *iMosflm* (Battye *et al.*, 2011) and scaled and merged using *AIMLESS* (Evans & Murshudov, 2013) from the *CCP*4 program suite (Winn *et al.*, 2011). Processed data are summarized in Table 3.



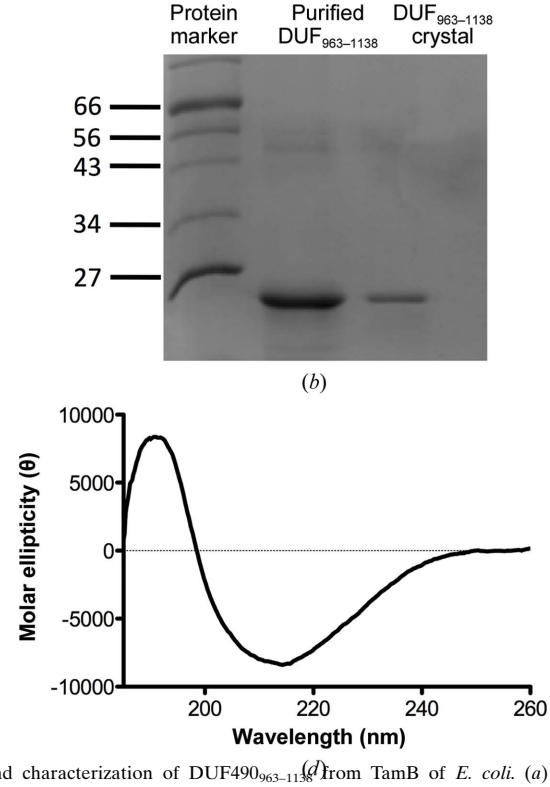


Figure 1
Purification and characterization of DUF490$_{963–1138}$ from TamB of *E. coli*. (*a*) Domain representation of TamB highlighting the C-terminal DUF490 domain and an N-terminal signal peptide that also serves as a membrane anchor. (*b*) 15% SDS–PAGE of purified DUF490$_{963–1138}$ shows a single band migrating at the approximate molecular weight expected for the construct (20 kDa). Protein markers are labelled in kDa. Additionally, several crystals were washed with 35% PEG 400 solution three times and then dissolved in pure H_2O and loaded onto the gel. (*c*) Secondary-structure prediction of the DUF490$_{963–1138}$ fold using *PSIPRED* suggests that the protein consists predominantly of β-strands and random coil. (*d*) The far-UV CD spectrum of the construct confirms that the protein consists predominantly of β-strands (30%), turns (25%) and disordered polypeptide (33%) with minor α-helical content (12%). Values were calculated based on the average of all matching solutions.

Table 1
Macromolecule-production information.

Source organism	*E. coli* strain K-12
Forward primer	TCAGCATATGATGGATGTATCGCCAGATGTTGTA
Reverse primer	GGTACTCGAGCGACATCGCCGGGTCAGA
Cloning vector	pET-21a
Expression host	*E. coli* BL21 (DE3)
Complete amino-acid sequence of the construct produced	MMDVSPDVVFEATPNLFTLDGRVDVPWARIVVHDLPESAVGV-SSDVVMLNDNLQPEEPKTASIPINSNLIVHVGNNVRIDAF-GLKARLTGDLNVVQDKQGLGLNGQINIPEGRFHAYGQDLI-VRKGELLFSGPPDQPYLNIEAIRNPDATEDDVIAGVRVTG-LADEPKAEIFSDPAMSLEHHHHHH

Table 2
Crystallization.

Method	Sitting-drop vapour diffusion
Plate type	96-well MRC plate
Temperature (K)	289
Protein concentration (mg ml^{-1})	15
Buffer composition of protein solution	50 m*M* Tris–HCl, 200 m*M* NaCl, 0.05% LDAO pH 7.5
Composition of reservoir solution	0.1 *M* HEPES, 15%(*v*/*v*) PEG 400, 0.2 *M* $CaCl_2$ pH 7.0
Volume and ratio of drop	1 µl, 1:1
Volume of reservoir (µl)	60

3. Results and discussion

DUF490$_{963–1138}$ from TamB was produced recombinantly and purified to homogeneity. The protein migrates as a single band on an SDS–PAGE gel, with an estimated molecular weight of 20 kDa, which is close to the calculated molecular weight based on the amino-acid sequence of the His$_6$-tagged protein (20.3 kDa; Fig. 1*b*). We observed the presence of multimeric species following gel-filtration chromatography (Supplementary Fig. S2). Whether this oligomerization is physiologically relevant is not currently known. However, in the previous study describing the discovery of the TAM complex, TamB was shown to behave as a monomer on blue native PAGE (Selkrig *et al.*, 2012). For crystallization purposes and further sample characterization we selected the gel-filtration peak that corresponded to the monomeric fraction. Far-UV circular-dichroism analysis of DUF490$_{963–1138}$ suggests that the protein consists predominantly of β-strands, turns and random coil, with a small α-helical fraction (Fig. 1*d*). This is in agreement with the analysis of the amino-acid sequence by the secondary-structure prediction program *PSIPRED* (McGuffin *et al.*, 2000; Fig. 1*c*).

Initial crystal screening consisted of 384 crystallization conditions, of which only one produced a successful hit. Optimization of this crystallization condition was undertaken varying the levels of precipitant, pH and protein concentration (Fig. 2). The presence of DUF490$_{963–1138}$ within crystals was confirmed by analysing washed crystals by SDS–PAGE (Fig. 1*b*). X-ray diffraction data were collected to a resolution of 2.1 Å (Fig. 3). The crystals belonged to space group $P3_121$ (or $P3_221$). The unit-cell parameters were $a = b = 57.34$, $c = 220.74$ Å, $\alpha = \beta = 90$, $\gamma = 120°$. Two monomers are predicted in the asymmetric unit, with a calculated Matthews coefficient of 2.90 Å^3 Da^{-1} and a solvent content of 57.6%. Analysis of possible twinning was carried out using *phenix.xtriage*, which detected no twinning (Zwart *et al.*, 2005). Owing to the absence of any structural models of DUF490$_{963–1138}$ homologues, *ab initio* modelling of DUF490$_{963–1138}$ was attempted using the *I-TASSER*, *Phyre* and *SWISS-MODEL* servers (Roy *et al.*, 2010; Kelley & Sternberg, 2009; Biasini *et al.*, 2014). However, the confidence in the models was very low and molecular replacement was unsuccessful. Initial experimental phasing with heavy-atom derivatives was also undertaken; however, heavy-metal soaks proved detrimental to the crystal quality

Table 3
Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	I03, DLS
Wavelength (Å)	0.9763
Temperature (K)	100
Detector	PILATUS3 6M
Rotation range per image (°)	0.15
Exposure time per image (s)	0.08
Space group	$P3_121$ or $P3_221$
a, b, c (Å)	57.34, 57.34, 220.74
α, β, γ (°)	90, 90, 120
Multiplicity	9.6 (9.5)
Resolution range (Å)	48.45–2.10 (2.17–2.10)
No. of unique reflections	25570 (2303)
Completeness (%)	99.7 (98.9)
$\langle I/\sigma(I)\rangle$	12.1 (3.2)
$R_{\rm p.i.m.}$† (%)	3.2 (26.2)
$R_{\rm merge}$‡ (%)	9.3 (76.3)

† $R_{\rm p.i.m.} = \sum_{hkl}\{1/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$. ‡ $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$.

Figure 2
Optimized crystals of DUF490$_{963–1138}$ grown using the sitting-drop vapour-diffusion method. The dimensions of the crystals were approximately 100 × 50 × 30 µm.

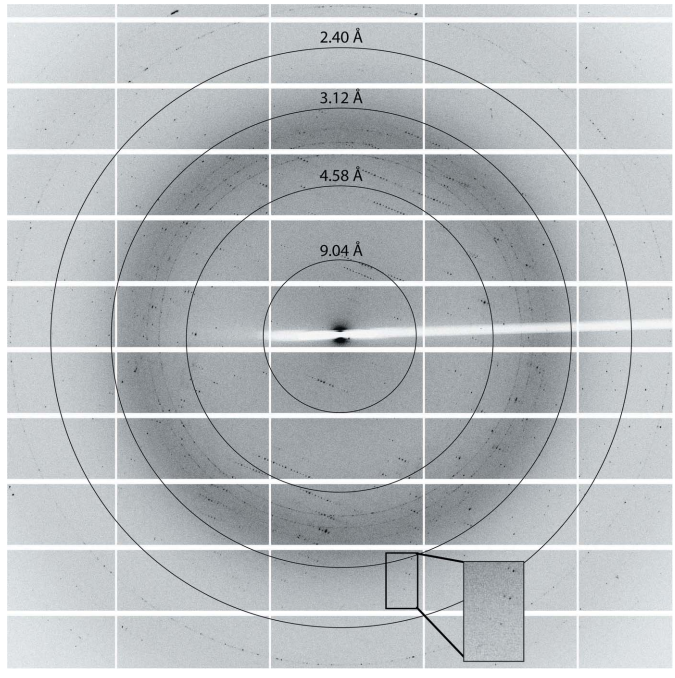


Figure 3
Representative diffraction pattern of DUF490$_{963–1138}$. The box highlights the high-resolution spots with an adjusted background.

and diffraction and co-crystallization trials failed to produce any crystals. Therefore, selenomethionine-labelled protein will be produced for experimental phasing (the hypothetical number of Met residues in the asymmetric unit is eight). The high-resolution structure of $DUF490_{963–1138}$ from TamB will provide insight into the function of this domain and its possible contribution to auto-transporter biogenesis.

We would like to acknowledge the Diamond Light Source for access to I03 (proposal No. MX5689). IJ is supported by a four-year studentship from the Wellcome Trust (grant No. 093597/Z/10/Z). RG is supported by a Kelvin Smith Scholarship from the University of Glasgow.

References

Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M. & Bordoli, L. (2014). *Nucleic Acids Res.* **42**, W252–W258.

Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.

Geer, L. Y., Domrachev, M., Lipman, D. J. & Bryant, S. H. (2002). *Genome Res.* **12**, 1619–1623.

Kelley, L. A. & Sternberg, M. J. (2009). *Nature Protoc.* **4**, 363–371.

McGuffin, L. J., Bryson, K. & Jones, D. T. (2000). *Bioinformatics*, **16**, 404–405.

Nishimura, K., Tajima, N., Yoon, Y.-H., Park, S.-Y. & Tame, J. R. H. (2010). *J. Mol. Med.* **88**, 451–458.

Provencher, S. W. & Glöckner, J. (1981). *Biochemistry*, **20**, 33–37.

Roy, A., Kucukural, A. & Zhang, Y. (2010). *Nature Protoc.* **5**, 725–738.

Sauri, A., Soprova, Z., Wickström, D., de Gier, J.-W., Van der Schors, R. C., Smit, A. B., Jong, W. S. P. & Luirink, J. (2009). *Microbiology*, **155**, 3982–3991.

Selkrig, J. *et al.* (2012). *Nature Struct. Mol. Biol.* **19**, 506–510.

Whitmore, L. & Wallace, B. A. (2008). *Biopolymers*, **89**, 392–400.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.

Zwart, P., Grosse-Kunstleve, R. & Adams, P. (2005). *CCP4 Newsl. Protein Crystallogr.* **43**, 27–35.